DMT of Multi-hop Cooperative Networks - Part II: Half-Duplex Networks with Full-Duplex Performance

K. Sreeram, S. Birenjith, and P. Vijay Kumar

***Abstract*— We consider single-source single-sink (ss-ss) multi-hop relay networks, with slow-fading links and single-antenna half-duplex relay nodes. While two-hop cooperative relay networks have been studied in great detail in terms of the diversity-multiplexing gain tradeoff (DMT), few results are available for more general networks. In a companion paper, we characterized end points of DMT of arbitrary networks, and established some basic results which laid the foundation for the results presented here. In the present paper, we identify two families of networks that are multi-hop generalizations of the two-hop network: K-Parallel-Path (KPP) networks and layered networks.**

KPP networks may be viewed as the union of K node-disjoint parallel relaying paths. Generalizations of these networks include KPP(I) networks, which permit interference between paths and KPP(D) networks, which possess a direct link between source and sink. We characterize the DMT of these families of networks completely for $K > 3$ and show that they can achieve the cut-set bound, thus proving that the DMT performance of full-duplex networks can be obtained even in the presence of the half-duplex constraint. We then consider layered networks, which are comprised of layers of relays, and prove that a linear DMT between the maximum diversity $d_{\max}$ and the maximum multiplexing gain of 1 is achievable for single-antenna fully-connected(fc) layered networks. This is shown to be equal to the cut-set bound on DMT if the number of relaying layers is less than 4, thus characterizing the DMT of this family of networks completely. For multiple-antenna KPP and layered networks, we provide lower bounds on DMT, that are significantly better than the best-known bounds.

All protocols in this paper are explicit and use only amplify-and-forward (AF) relaying. We also construct codes with short block-lengths based on cyclic division algebras that achieve the optimal DMT for all the proposed schemes. In addition, it is shown that codes achieving full diversity on a MIMO Rayleigh channel achieve full diversity on arbitrary fading channels as well.

Two key implications of the results in the paper are that the half-duplex constraint does not entail any rate loss for a large class of cooperative networks and that simple AF protocols are often sufficient to attain the optimal DMT.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India (Email: sreeramkannan@ece.iisc.ernet.in, biren@ece.iisc.ernet.in, vijayk@usc.edu). P. Vijay Kumar is on leave of absence from the University of Southern California, Los Angeles, USA.

This work was supported in part by NSF-ITR Grant CCR-0326628, in part by the DRDO-IISc Program on Advanced Mathematical Engineering and in part by Motorola's University Research Partnership Program with IISc.

The material in this paper was presented in part at the 10th International Symposium on Wireless Personal Multimedia Communications, Jaipur, Dec. 2007, the Information Theory and Applications Workshop, San Diego, Jan. 2008, and at the IEEE International Symposium on Information Theory, Toronto, July 2008.

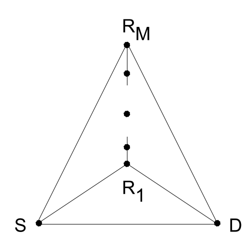
I. INTRODUCTION

In fading relay networks, cooperative diversity provides a method of efficient operation of networks. While much of work in the literature on cooperative diversity is based on two-hop networks, we focus our attention on multi-hop networks. For a review of related literature, see Section I-A in the companion paper [1]. In the companion paper, we derived results pertaining to the DMT of arbitrary full-duplex networks.

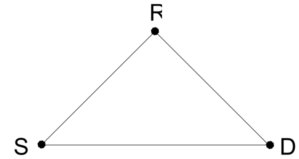
In the present paper, we deal with half-duplex networks, for which specification of explicit schedules requires some structure in the network. Therefore, we focus on specific classes of half-duplex networks in the present paper. For half-duplex networks without a direct link, even the achievability of a maximum multiplexing gain of 1, in the case of single antenna networks, is not clear. We will show that for a large family of networks, this maximum multiplexing gain can be achieved by appropriately scheduling the links. In fact, we will show that the cut-set upper bound on DMT for many of these networks can be achieved, thus demonstrating that the half-duplex operation does not entail any loss in DMT performance as compared to full-duplex operation for these families of networks.

A. Classification of Networks

In this section, we define the classes of networks under consideration. The well-studied two-hop network with direct link is shown in Fig. 1. We will consider two multi-hop generalizations of two-hop networks in this paper: KPP and layered networks. Unless otherwise stated, all networks considered possess a single source and a single sink and we will apply the abbreviation ss-ss to these networks.



(a) General two-hop relay network



(b) Single relay network

Fig. 1. Two-hop cooperative relay networks

A cooperative wireless network can be built out of a collection of spatially distributed nodes in many ways. For instance, we can identify paths connecting source to the sink through a series of nodes in such a manner that any two adjacent nodes fall within the Rayleigh zone [3]. This process can be continued barring those nodes which are already chosen. Such a construction will result in a set of paths from the source to the sink. In the simplest model, we can further impose the constraint that these paths do not interfere each other, see Fig.1(a), thus motivating the study of a class of multi-hop network which we shall refer to as the set of K-Parallel-Path (KPP) networks.

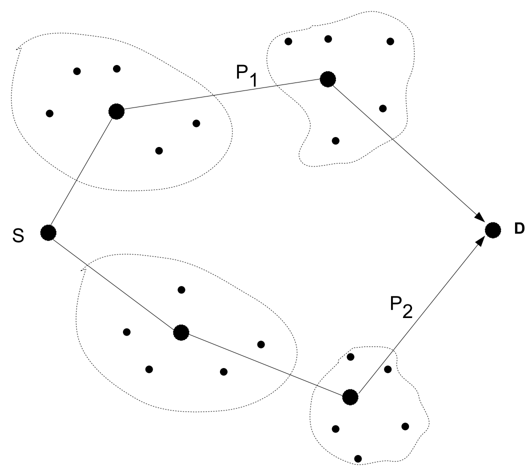


Fig. 2. Motivation for the KPP networks

Alternatively, layers of relays can be identified from a collection of nodes between the source and the sink. This will result in a layered network model.

1) Network Representation by Graph: Any wireless network can be associated with a directed graph, with vertices representing nodes in the network and edges representing connectivity between nodes. If an edge is bidirectional, we will represent it by two edges one pointing in either direction. An edge in a directed graph is said to be *live* at a particular time instant if the node at the head of the edge is transmitting at that instant. An edge in a directed graph is said to be *active* at a particular time instant if the node at the head of the edge is transmitting and the tail of the edge is receiving at that instant. Since most networks considered in this paper have bidirectional links, we will represent a bidirectional link by an undirected edge. Therefore, undirected edges must be interpreted as two directed edges, with one edge pointing in either direction.

A wireless network is characterized by broadcast and interference constraints. Under the *broadcast* constraint, all edges connected to a transmitting node are simultaneously live and transmit the same information. Under the *interference* constraint, the symbol received by a receiving end is equal to the sum of the symbols transmitted on all incoming live edges. We say a protocol avoids interference if only one incoming edge is live for all receiving nodes.

In wireless networks, the relay nodes operate in either half or full-duplex mode. In case of half-duplex operation, a node cannot simultaneously listen and transmit, i.e., an incoming edge and an outgoing edge of a node cannot be simultaneously active.

2) K-Parallel-Path Networks: One way of generalizing the two-hop relay network is to consider this network as a collection of K parallel, relaying paths from the source to the sink, each of length greater than 1. This immediately leads to a more general network that is comprised of K parallel paths of varying lengths, linking source and sink. More formally:

Definition 1: A set of edges (v_1, v_2), (v_2, v_3), $\ldots$, (v_{n-1}, v_n) connecting the vertices v_1 to v_n is called a path. The length of a path is the number of edges in the path. The K-parallel path (KPP) network is defined as a ss-ss network that can be expressed as the union of K node-disjoint paths, each of length greater than one, connecting the source to the sink. Each of the node-disjoint paths is called a relaying path. All edges in a KPP network are bidirectional (see Fig. 3).

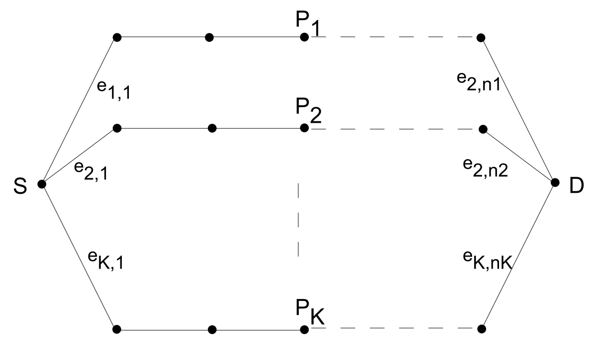


Fig. 3. The KPP network

The communication between the source and the sink takes place along K parallel paths, labeled with the indices P_1, P_2, $\ldots$, P_K. Along path P_i, the information is transmitted from source to sink through multiple hops with the aid of $n_i - 1$ intermediate relay nodes $\{R_{ij}\}_{j=1}^{n_i-1}$.

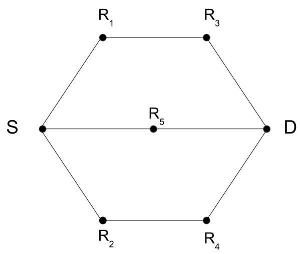
Remark 1: A network similar to the KPP network in Definition 1 is considered in [26], albeit from a symbol error probability perspective.

Definition 1 of KPP networks precludes the possibility of either having a direct link between the source and the sink, or of having links connecting nodes lying on different node-disjoint paths. We now extend the definition of KPP networks to include these possibilities.

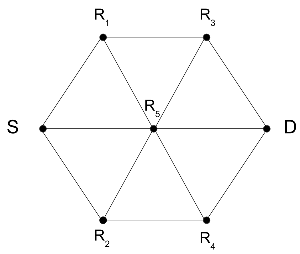
Definition 2: If a given network is a union of a KPP network and a direct link between the source and sink, then the network is called a KPP network with direct link, denoted by KPP(D). If a given network is a union of a KPP network and links interconnecting relays in various paths, then the network is called a KPP network with interference, denoted by KPP(I). If a given network is a union of a KPP network, a direct link and links interconnecting relays in various paths, then the network is called a KPP network with interference and direct path, denoted by KPP(I, D).

Fig. 4 below provides examples of all four variants of KPP networks.

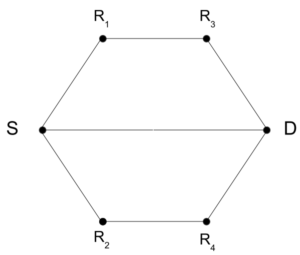
For a KPP(D), KPP(I) or KPP(I, D) network, we consider the union of the K node-disjoint paths as the backbone KPP



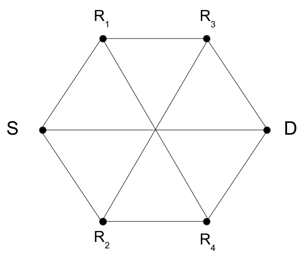
(a) A KPP network with $K = 3$



(b) A KPP(I) network with $K = 3$



(c) A KPP(D) network with $K = 2$



(d) A KPP(I, D) network with $K = 2$

Fig. 4. Examples of KPP networks.

network. While there may be many choices for the K node-disjoint paths, we can choose any one such choice and call that the backbone KPP network. These K relaying paths in these networks are referred to as the K *backbone* paths. A *start node* and *end node* of a backbone path are the first and the last relays respectively in the path. There are precisely K start nodes in a KPP network, which are connected at one end to the source. This remains the case even for KPP(I) and KPP(I, D) networks. Similarly, sink node is connected to exactly K end nodes in KPP, KPP(I) and KPP(I, D) networks.

In a general KPP network, let $P_i, i = 1, 2, ..., K$ be the K backbone paths. Let P_i have n_i edges. The j-th edge on the i-th path P_i will be denoted by e_{ij} and the associated fading coefficient by g_{ij}.

3) Layered Networks: A second way of generalizing a two-hop relay network is to view the two-hop network as a network comprising of a single layer of relays. The immediate generalization is to allow for more layers of relays between source and sink, with the proviso that any link is either between nodes in adjacent layers or connects two nodes in the same layer. We label this class of multi-hop relaying networks as *layered networks*:

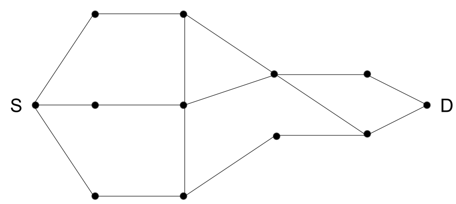
Definition 3: Consider a ss-ss single-antenna bidirectional network. A network is said to be a layered network if there exists a partition of the vertex set V into subsets $V_0, V_1, ..., V_L, V_{L+1}$ such that

- V_0, V_{L+1} denote the singleton sets corresponding to the source and sink respectively and for all $0 < i < L+1$, $|V_i| \geq 2$.
- If there is an edge between a node in vertex set V_i and a node in V_j, then $|i - j| \leq 1$. We assume $|V_i| > 1, i = 1, 2, .., L$

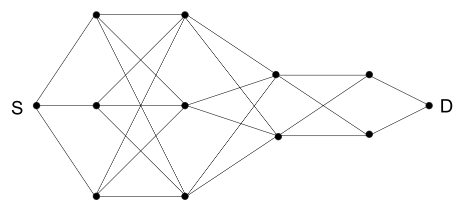
We will refer to $V_1, ..., V_L$ as the relaying layers of the network. A layered network is said to be fully-connected (fc) if for any i, $v_1 \in V_i$ and $v_2 \in V_{i+1}$, (v_1, v_2) is an edge in the network. For fc layered networks, we include an additional condition that the relaying layers have at least two relays in each layer, i.e., $V_i > 1, i = 1, 2, ..., L$.

It must be noted that a fc layered network may or may not have links within a layer. Therefore, whenever we say fc layered network, it applies to both networks that have intra-layer links and those that do not have such links. Examples of both these types of networks are shown in Fig. 5(b) and Fig. 5(c).

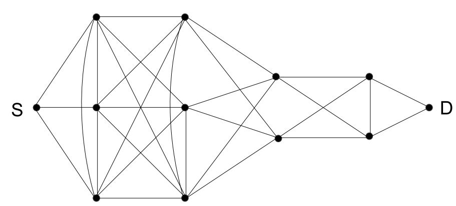
Remark 2: The definition of a layered network is general enough to accommodate all ss-ss networks without a direct link. This is because, any ss-ss network can be re-drawn as a layered network with a single layer comprising of all relays in the network and interconnections between relays. For this general case, we give a certain achievable DMT.



(a) A layered network with with 4 relaying layers



(b) A fully-connected layered network



(c) A fully-connected layered network with intra-layer links

Fig. 5. Examples of layered networks

Every layered network will have a layer containing only the source, and a second layer containing only the sink. In Fig.5, examples of layered networks are given. Layered networks were also considered in [18], [9] and [21]. In particular, [18] considered layered networks having an equal number of relay nodes in all layers. We will refer to such layered networks as regular networks and we will formally define them below.

Remark 3: In this remark, we characterize the intersection of KPP(I) networks and layered networks. First we observe that one is not contained in the other. Consider the subgraph of a given KPP(I) network graph, consisting of all the nodes of the original network except for the source and the sink. This subgraph will have the property that the number of node-disjoint and edge-disjoint paths is equal to the number of relay nodes immediately adjacent to the source. This is a key property of KPP(I) networks, which in general, does not hold for layered networks. On the other hand, there

can be cross links between the parallel paths in a KPP(I) network in such a way that the network cannot be viewed as being layered. However, these two classes of networks are not mutually exclusive and in fact, we term networks that lie in the intersection of the two classes as regular networks.

Definition 4: A (K, L) Regular network is defined as a KPP(I) network that is also, simultaneously, a layered network with L layers of relays (see Fig. 6).

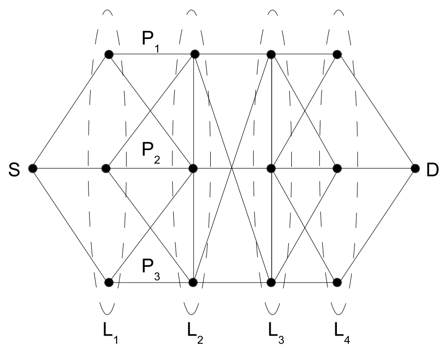


Fig. 6. A regular network with 4 layers and 3 paths

Remark 4: The two-hop relay network [Fig.1(a)] is a KPP(I,D) network with $K = M$, M being the number of relays. In the absence of a direct link, the two-hop relay network is a KPP(I) network with $K = M$. In fact, the two-hop relay network without direct link is also a layered network with a single layer of relays, thereby making it a $(M, 1)$ regular network. On the other hand, if we have a two-hop relay network with direct link but make the additional assumption of relay isolation, then it is a KPP(D) network with $K = M$.

B. Setting and Channel Model

We use uppercase letters to denote matrices and lowercase letters to denote vectors/scalars. Vectors and scalars are differentiated between each other by the context. Irrespective of whether it is a scalar, vector or a matrix, boldface letters are used to denote random entities.

Between any two adjacent nodes v_x and v_y of a wireless network, we assume the following channel model.

$$\mathbf{y} \ = \mathbf{H}\mathbf{x} + \mathbf{w} \ , \tag{1}$$

where $\mathbf{y}$ corresponds to the received signal at node v_y, $\mathbf{w}$ is the noise vector, $\mathbf{H}$ is the channel matrix and $\mathbf{x}$ is the vector transmitted by the node v_x.

1) Assumptions: We follow the literature in making the assumptions listed below. Our description is in terms of the equivalent complex-baseband, discrete-time channel.

1) All channels are assumed to be quasi-static and to experience Rayleigh fading and hence all fade coefficients are i.i.d., circularly-symmetric complex Gaussian $\mathbb{C}\mathcal{N}(0, 1)$ random variables.
2) The additive noise at each receiver is also modeled as possessing an i.i.d., circularly-symmetric complex Gaussian $\mathbb{C}\mathcal{N}(0, 1)$ distribution.
3) Each receiver (but none of the transmitters) is assumed to have perfect channel state information of all the upstream channels in the network. [1]

C. Background

We refer the reader to Section II-A.1 of the companion paper [1] for a background on the diversity-multiplexing gain tradeoff (DMT).

1) Cut-set Bound on DMT: For each of the networks described in this paper, we can get an upper bound on the DMT, based on the cut-set upper bound on mutual information [25]. This was formalized in [3] as follows:

Lemma 1.1: Let r be the rate of multiplexing gain at which communication between the source and the sink is taking place. Given a cut ω, there is a channel matrix $\mathbf{H}_\omega$ connecting the input terminals of the cut to the output terminals. Let us call the DMT corresponding to this $\mathbf{H}_\omega$ matrix as the DMT of the cut, $d_\omega(r)$. Then the DMT between the source and the sink is upper bounded by

$$d(r) \leq \min_{\omega \in \Lambda} \{d_\omega(r)\} \ ,$$

where Λ is the set of all cuts between the source and destination.

Remark 5: Note that the cut-set bound does not take into account half-duplex operation of the network and therefore applies equally to both full and half-duplex networks. This clearly presents a greater challenge for half-duplex protocols.

Definition 5: Given a random matrix $\mathbf{H}$ of size $m \times n$, we define the *DMT of the matrix* $\mathbf{H}$ as the DMT of the associated channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ where $\mathbf{y}$ is a m length received column vector, $\mathbf{x}$ is a n length transmitted column vector and $\mathbf{w}$ is a $\mathbb{C}\mathcal{N}(0, I)$ column vector. We denote the DMT by $d_H(.)$

2) Amplify and Forward Protocols : [2]

An AF protocol $\wp$ is a protocol $\wp$ in which each node in the network operates in an amplify-and-forward fashion. Such protocols induce a linear channel model between source and sink of the form:

$$\mathbf{y} \ = \mathbf{H}(\wp)\mathbf{x} + \mathbf{w} \ , \tag{2}$$

where $\mathbf{y} \in \mathbb{C}^m$ denotes the signal received at the sink, $\mathbf{w}$ is the noise vector, $\mathbf{H}(\wp)$ is the $(m \times n)$ induced channel matrix and $\mathbf{x} \in \mathbb{C}^n$ is the vector transmitted by the source. We impose the following energy constraint on the transmitted vector $\mathbf{x}$

$$\mathrm{Tr}(\Sigma_x) \ := \ \mathrm{Tr}(\mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}) \quad \leq \quad n\rho$$

where Tr denote the trace operator, and we will regard ρ as representing the SNR on the network. We will assume a symmetric power constraint on the relays and the source. However it will turn out that given our high SNR perspective, the exact power constraint is not very important. We consider both half and full-duplex operation at the relay nodes.

[1]However, for the protocols proposed in this paper, the CSIR is utilized only at the sink, since all the relay nodes are required to simply amplify and forward the received signal.

[2]This section is the same as Section II-A.3 in the first part of the paper [1] and is included here for ease of reference.

Our attention here will be restricted to amplify-and-forward (AF) protocols since as we shall see, this class of protocols can often achieve the DMT of a network. More specifically, our protocol will require the links in the network to operate according to a schedule which determines the time slots during which a node listens as well as the time slots during which it transmits. When we say that a node listens, we will mean that the node stores the corresponding received signal in its buffer. When a node does transmit, the transmitted signal is simply a scaled version of the most recent received signal contained in its buffer, with the scaling constant chosen to meet a transmit power constraint. [3] In particular, nodes in the network are not required to decode and then re-encode. It turns out [2], that the value of the scaling constant does not affect the DMT of the network operating under the specific AF protocol. Without loss of accuracy therefore, we will assume that this constant is equal to 1.

It follows that, for any given network, we only need specify the schedule to completely specify the protocol. This will create a virtual MIMO channel of the form $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ where $\mathbf{H}$ is the effective transfer matrix and $\mathbf{w}$ is the noise vector, which is in general colored.

D. Certain Results from the Companion Paper

In the companion paper [1], we developed basic results that will be instrumental in deriving the DMT of certain classes of networks in this paper. A few important results among them are given here for reference.

We proved that the correlated noise encountered at the sink of many multi-hop networks is white in the scale of interest. This result will be assumed throughout this paper and is formalized in the theorem below.

Theorem 1.2: [1] Consider a channel of the form $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{z}$. Let $\mathbf{h_1}, \mathbf{h_2}, ..., \mathbf{h_L}$ be L, possibly dependent, Rayleigh random variables. Let $\mathbf{G}_i, i = 1, 2, .., M$ be $N \times N$ matrices in which each entry is a polynomial function of the random variables $\mathbf{h_1}, \mathbf{h_2}, ..., \mathbf{h_L}$. Let $\mathbf{z} = \mathbf{z}_0 + \sum_{i=1}^{M} \mathbf{G}_i\mathbf{z}_i$ be the noise vector. Let $\{\mathbf{z}_i\}$ be i.i.d. circularly symmetric, n-dimensional complex gaussian $\mathbb{C}\mathcal{N}(\underline{0}, I)$ random vectors. The random matrix $\mathbf{H}$ in general depends on the random variables $\mathbf{h_i}$. Then $\mathbf{z}$ is white in the scale of interest, i.e.,

$$\begin{aligned} &Pr(\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}\Sigma^{-1}) \leq r\log\rho) \\ \doteq\quad &Pr(\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}) \leq r\log\rho) \end{aligned}$$

We also proved a result pertaining to the DMT of block lower triangular(blt) matrices as given below.

Theorem 1.3: [1] Consider a random blt matrix $\mathbf{H}$ having component matrices $\mathbf{H}_{ij}$ of size $N_i \times N_j$. Let $M := \sum_{i=1}^{N} N_i$ be the size of the square matrix $\mathbf{H}$.

Let $\mathbf{H}^{(0)}$ be the diagonal part of the matrix $\mathbf{H}$ and $\mathbf{H}^{(\ell)}$ denote the last sub-diagonal matrix of $\mathbf{H}$. Then,

1) $d_H(r) \geq d_{H^{(0)}}(r)$.
2) $d_H(r) \geq d_{H^{(\ell)}}(r)$.
3) In addition, if the entries of $H^{(\ell)}$ are independent of the entries in $H^{(0)}$, then $d_H(r) \geq d_{H^{(0)}}(r) + d_{H^{(\ell)}}(r)$

[3] More sophisticated linear processing techniques would include matrix transformations of the incoming signal, but turns out to be not needed here.

In this paper, we will frequently use two results on the DMT of parallel channel that are proved in the companion paper [1].

Lemma 1.4: [1] Consider a parallel channel with M links, with the ith link having representation $\mathbf{y_i} = \mathbf{H_i}\mathbf{x_i} + \mathbf{w_i}$, and let $d_i(\cdot)$ denote the corresponding DMT. Then the DMT of the overall parallel channel is given by

$$d(r) = \inf_{(r_1, r_2, \cdots, r_M):\ \sum_{i=1}^{M} r_i = r} \sum_{i=1}^{M} d_i(r_i). \tag{3}$$

Lemma 1.5: [1] Consider a parallel channel with M links and repeated channel matrices. More precisely, let there be N distinct channel matrices $H^{(1)}, H^{(2)}, ..., H^{(N)}$, with $H^{(i)}$ repeating in n_i sub-channels, such that $\sum_{i=1}^{N} n_i = M$. Then the DMT of such a parallel channel is given by,

$$d(r) = \inf_{(r_1, r_2, \cdots, r_M):\ \sum_{i=1}^{N}\ n_i r_i = r} \sum_{i=1}^{N} d_i(r_i). \tag{4}$$

We also established that diversity of any flow in a multi-terminal network equals the min-cut between the source and the sink of that flow. The result is given in the below theorem.

Theorem 1.6: [1] Consider a multi-terminal fading network with nodes having multiple antennas with edges connecting antennas on two different nodes having i.i.d. Rayleigh-fading coefficients. The maximum diversity achievable for any flow is equal to the min-cut between the source of the flow and the corresponding sink. Each flow can achieve its maximum diversity simultaneously.

We established an achievable DMT region for full-duplex networks, which is summarized in the following theorem.

Theorem 1.7: [1] Consider a ss-ss full-duplex network with single antenna nodes. Let the min-cut of the network be M. Let the network satisfy *either* of the two conditions below:

1) The network has no directed cycles, or
2) There exist a set of M edge-disjoint paths between source and sink such that *none* of the M paths have shortcuts.

Then, a linear DMT $d(r) = M(1 - r)^+$ between a maximum multiplexing gain of 1 and maximum diversity M is achievable.

E. Results

In this paper, we characterize the DMT of KPP networks and its variants called KPP(D) and KPP(I) networks. We also provide an achievable DMT for layered networks. In many cases, the achievable DMT equals the cut-set bound and is thereby optimal. All the strategies are of half-duplex nature. We give explicit protocols and code constructions for all cases. Some of these results were presented in conference versions of this paper [12]–[15] (see also [16], [17]).

The principal results of the paper are the following(see Table I).

1) For KPP, KPP(I) and KPP(D) networks, we propose an explicit protocol whose achievable DMT coincides with the cut-set bound.
2) For fc layered networks, we construct protocols that achieve a DMT that is linear between the maximum

diversity and maximum multiplexing gain points. This DMT is optimal if the number of layers is strictly less than 4.
3) For general layered networks, we give a sufficient condition for the achievability of a linear DMT between the maximum diversity and the maximum multiplexing gain in Lemma 4.3.
4) For KPP and layered networks with multiple antenna nodes, we examine certain protocols and establish achievable DMT for these protocols.
5) In Section VI, we give explicit codes with short block-lengths based on cyclic division algebras that achieve the best possible DMT for all the schemes proposed above. We also prove that full diversity codes for all networks in this paper can be obtained by using codes that give full diversity on a Rayleigh fading MIMO channel.

F. Outline

In Section II, we focus on half-duplex KPP networks and present protocols achieving optimal DMT for $K \geq 3$. We extend this result to KPP(D) networks at the end of this section. In Section III, KPP(I) networks with half-duplex relays are considered, and schemes achieving optimal DMT are presented for KPP(I) networks allowing certain types of interference. In Section IV, we consider layered networks and show that a linear DMT between maximum multiplexing gain and maximum diversity is achievable. In Section V, we consider multi-antenna layered and KPP networks and give an achievable DMT. Finally, in Section VI, we give explicit CDA based codes of low complexity for all the DMT optimal protocols.

II. Half-Duplex Networks with Isolated Paths - KPP Networks

In this section, we consider KPP and KPP(D) networks with single-antenna nodes operating under the half-duplex constraint. Discussion on KPP(I) nertworks is deferred to a later section.

1) The Cut-Set Bound for KPP Networks: The cut-set (Lemma 1.1) upper bound on DMT for the class of KPP networks is given by:

$$d(r) \ \leq \ K(1-r)^{+},$$

and our aim is to design protocols that attain this cut-set bound.

In this section we will present protocols for KPP networks, that permit the cut-set bound to be attained for all $K \geq 3$. For the case $K = 2$, we present a lower bound to the DMT that can be achieved using an AF protocol.

A. Full-Duplex KPP Networks

To start with, we will consider full-duplex KPP networks and review the results of the first part of this paper [1] as applied to KPP networks. Theorem 1.7 gives a lower-bound on the DMT of certain classes of networks with full-duplex relays. The result applies to KPP and KPP(I) networks, even with multiple antenna nodes, and is given in the corollary below.

Corollary 2.1: For the full-duplex KPP networks without direct link (i.e. KPP and KPP(I) networks), with potentially multiple antenna nodes, a DMT of $d_{\max}(1-r)^{+}$ is achievable.

When nodes have a single antenna, $K(1-r)^{+}$ turns out to be the optimal DMT. However, the characterization of optimal DMT becomes tricky when the relays are half-duplex. Most articles in the literature deal with half-duplex networks by first analyzing the networks from a full-duplex perspective and then translating the results by conceding rate loss by a factor of 2. In this and the next section, we will show that half-duplex multi-hop networks (specifically KPP networks) operating under a suitable schedule can achieve the same DMT performance as full-duplex networks.

B. Protocols Achieving the Cut-Set Bound for $K \geq 3$

As made clear in the introduction, we restrict our attention in this paper to the class of AF protocols under which node operations are restricted to scaling and forwarding. To completely specify the manner in which the network is operated, it remains only to identify a schedule of operation.

In all of the schedules considered here, node operations are periodic with a period of N time slots. Thus all edge activations are periodic as well, and we will refer to N as the cycle length of the protocol. We shall describe all our protocols in a simple manner, as an edge coloring scheme. Let $C = \{c_1, c_2, \ldots, c_N\}$ be the set of N colors used in the scheme. Each color in C signifies a distinct time slot within the N-slot protocol. There is a natural order between any two colors, inherited from the time-slots they represent within a cycle. All the edges in the network are assigned a subset of colors from the set C. The subset of colors assigned to the edge e_{ij} will be denoted by A_{ij}. Each color in A_{ij} represents the time instants during which the edge e_{ij} is active. [4]

Definition 6: A half-duplex protocol is said to be an orthogonal protocol if at any node, at a given time instant, only one of the incoming or outgoing edges is active. An orthogonal protocol for a KPP network is said to have the equi-activation property if each edge along a given backbone path in the KPP network is activated an equal number of times.

Remark 6: The definition of orthogonal protocol is similar in spirit to the definition in the networking literature [24], where a network is said to have orthogonal channels if interference is avoided at all nodes and each node is permitted to communicate with at most one other node at any given time.

All of the orthogonal protocols employed in this paper will satisfy the equi-activation property and hence, whenever in the sequel we speak of an orthogonal protocol, we will mean an orthogonal protocol that in addition, satisfies the equi-activation property.

Proposition 1: Every orthogonal protocol can be described as an edge coloring of the network satisfying the following constraints. Conversely, every edge coloring satisfying the following constraints describes an orthogonal protocol.

[4] We assume that the network is in operation for sufficient amount of time, so that if an edge is active, the node at beginning of the edge always has a symbol to transmit.

TABLE I
PRINCIPAL RESULTS SUMMARY

Network	No of sources/ sinks	No of antennas in nodes	Direct Link	Upper bound on Diversity/DMT $d_{\text{bound}}(r)$	Achievable Diversity/DMT $d_{\text{achieved}}(r)$	Is upper bound achieved?	Reference
KPP(K $\geq$ 3)	Single	Single	$\times$	$K(1-r)^+$	$K(1-r)^+$	✓	Theorem 2.9
KPP(D)(K $\geq$ 4)	Single	Single	✓	$(K+1)(1-r)^+$	$(K+1)(1-r)^+$	✓	Corollary 2.11
KPP(I)(K $\geq$ 3)	Single	Single	$\times$	$K(1-r)^+$	$K(1-r)^+$	✓	Theorem 3.6
Fully Connected Layered	Single	Single	$\times$	Concave in general	$M(1-r)^+$	A linear DMT between $d_{\max}$ and r_{max} is achieved. ✓ for $L < 4$	Theorem 4.6 Corollary 4.7
Any network satisfying Lemma 4.3	Single	Single	$\times$	Cut-set Bound	$M(1-r)^+$	A linear DMT between $d_{\max}$ and r_{max} is achieved	Lemma 4.3
(K, L) Regular	Single	Single	$\times$	$K(1-r)^+$	$K(1-r)^+$	✓	Theorem 3.2

$$A_{i1} \cap A_{j1} = \phi, i \neq j. \tag{5}$$
$$A_{in_i} \cap A_{jn_j} = \phi, i \neq j. \tag{6}$$
$$A_{ij} \cap A_{i(j+1)} = \phi, j = 1, 2, ..., n_i - 1. \tag{7}$$
$$|A_{ij}| = m_i, j = 1, 2, ..., n_i. \tag{8}$$

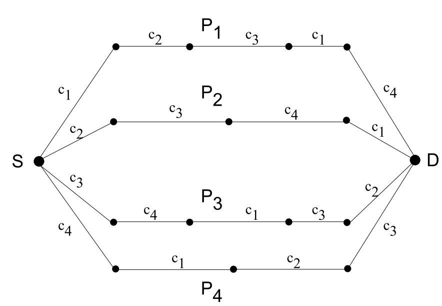


Fig. 7. KPP network orthogonal protocol

The first constraint corresponds to the fact that for an orthogonal protocol, only one outgoing edge is active at the source. Similarly the second constraint corresponds to the fact that for an orthogonal protocol, only one incoming edge is active at the sink. The third constraint captures the half-duplex nature of the protocol. The last constraint corresponds to the equi-activation property of the orthogonal protocols considered here. A coloring which respects all the above constraints for an example KPP network is given in Fig. 7. In this example, the protocol has a cycle length of 4 time-slots, represented by the colors c_1, c_2, c_3, and c_4 in order. Within a given cycle, each edge is activated during exactly one time-slot, specified by a color, c_i, $1 \leq i \leq 4$.

Definition 7: The rate, R of an orthogonal protocol is defined as the ratio of the number of symbols transmitted by the source to the total number of time slots. In terms of the notation above, we have

$$R = \frac{\sum_{i=1}^{K} m_i}{N}.$$

Definition 8: Consider a KPP network. Let v_1, v_2, v_3, v_4 be four consecutive vertices lying on one of the K parallel paths leading from the source to the sink. Let v_1 and v_3 transmit, thereby causing the edges (v_1, v_2) and (v_3, v_4) to be active. Due to the broadcast and interference constraints, transmission from v_3 interferes with the reception at v_2. This is termed as back-flow, and is illustrated in Fig.8

Back-flow can be avoided by ensuring that there are at least two inactive edges between any two active edges along a backbone path. We formalize this below:

Remark 7: An orthogonal protocol avoids back-flow if the

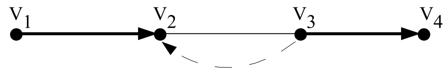


Fig. 8. Back-flow on a path

edge coloring satisfies:

$$A_{ij} \cap A_{i(j+2)} = \phi, j = 1, 2, ..., n_i - 2.$$

From the remark above, it is evident that any three adjacent edges $e_{ij}, e_{i(j+1)}$, and $e_{i(j+2)}$ will map to disjoint sets of colors when the coloring scheme corresponds to an orthogonal protocol that avoids back-flow. Moreover, repetition of the same set of colors along every third edge is permitted under the constraints of an orthogonal protocol avoiding back-flow. This suggests an easy way of arriving at an edge coloring that satisfies the constraints. To color the edges along a backbone path in the KPP network, we will first select three disjoint subsets of colors, order these and then repeat them cyclically along the edges of the path moving from source to sink. For reasons that will become apparent later, we may deviate from this rule with respect to the coloring of the last edge along any backbone path. More formally, for any given path $P_i, i \in [K]$, let $G_i = (G_{i0}, G_{i1}, G_{i2})$ be the set of three colors repeated cyclically starting from the first node and let F^i specify the subset assigned to the last edge e_{in_i} along path P_i. Then the corresponding edge coloring is given by

$$A_{ij} = \begin{cases} G_{i(j \pmod 3)}, & j \neq n_i \\ F^i, & j = n_i \end{cases}. \tag{9}$$

Henceforth, we will use G^i, F^i as descriptors of the orthogonal protocol. We show next how by adding a few additional constraints, we obtain a protocol that achieves the cut-set bound.

Lemma 2.2: Consider a KPP network. If an orthogonal protocol satisfies the following constraints:

1) The rate of the protocol is equal to one.
2) In every cycle, the sink receives an equal number of symbols from each of the K backbone paths.
3) The protocol avoids back-flow.

Then the protocol achieves the cut-set bound, i.e.,

$$d(r) = K(1-r)^+.$$

Proof: Consider a situation when a KPP network starts operating under an orthogonal protocol satisfying the constraints above. Since the sink receives an equal number of symbols from each of the K backbone paths, symbols from the source should be transmitted via each of the K backbone paths equally often. Also, the protocol being of rate 1, a symbol is transmitted from the source in every time-slot. For this reason, we treat the transmission from the source in blocks constituting of K symbols, each denoted by a vector $\mathbf{x} = [\mathbf{x}_1, \cdots, \mathbf{x}_K]^t$, in which a symbol $\mathbf{x}_i$ takes on backbone path P_i. Clearly, the delay encountered by each symbol depends on the length of the path it traverses. However, it also depends on the schedule of edge activations dictated by the protocol. For example, in a path P_1, consider two adjacent edges e_{1j} and $e_{1(j+1)}$ with a common node v. If e_{1j} and $e_{1(j+1)}$ are assigned with colors c_j and c_{j+2}(Here, we assume that time-slots are in the same order as that of index of the color), then v incurs a delay of 1 time-slot due to the schedule. Thus each component $\mathbf{x}_i$ experiences the sum of delays along all the nodes in the path P_i.

This difference in delay experienced by each component $\mathbf{x}_i$ in $\mathbf{x}$ may be large enough so that the information corresponding to $\mathbf{x}_i$ arrives at the sink node during one cycle of the protocol, whereas the information corresponding to $\mathbf{x}_{i+1}$ may arrive at the sink node during some other cycle. Therefore, there is a need to synchronize these information symbols so that they arrive at the same cycle of the protocol. This can be done by adjusting the delays of each path at the source (by a multiple of K) so that all the symbols $\mathbf{x}_1, ..., \mathbf{x}_K$ in a block of data are received in the same cycle at the destination. However, the activation of the edges connecting to the sink need not be in the same order as that in which edges connected to the source are activated. Therefore, within a cycle, the symbols might be received in a different order. Hence, we shuffle the input appropriately within each cycle (it corresponds to multiplication by a permutation matrix, P), so that we get an input output relation between the input $\mathbf{x}$ and the output $\mathbf{y}$ as,

$$\mathbf{y} = \begin{bmatrix} \mathbf{g}_1 & & & \\ & \mathbf{g}_2 & & \\ & & \ddots & \\ & & & \mathbf{g}_K \end{bmatrix} \mathbf{x} + \mathbf{w}, \tag{10}$$

where $\mathbf{g}_i$ denotes the product fading coefficient of the path P_i, and $\mathbf{w}$ is the noise vector at the sink, which is white in the scale of interest.

Note that the initial delay, D, encountered in the transmission is not accounted for in the channel model given in (10). However, when the network operates continuously for long duration T, this delay does not affect in terms of loss in rate since the rate loss factor of $\frac{T-D}{T}$ can be made arbitrarily small when T becomes large. We will always make this assumption whenever we are dealing with KPP, KPP(I), KPP(D) networks.

The DMT of the above matrix, $\mathbf{H}$ can be easily computed to be equal to

$$d(r) = K(1-r)^+. \tag{11}$$

■

Remark 8: In later sections, whenever we refer to orthogonal protocols for KPP networks satisfying conditions in Lemma 2.2, we assume that the protocol will induce a channel as given in (8), neglecting the initial delay. This is justified in the proof of the above lemma.

We will now show that the requirements of Lemma 2.2 can be met in the case of KPP networks with $K \geq 4$:

Theorem 2.3: For any KPP network with $K \geq 4$, there exists an explicit protocol achieving the cut-set bound on DMT.

Proof: We now establish an orthogonal protocol satisfying the conditions in Lemma 2.2 for the case when $K \geq 4$. We choose the cycle length of the protocol to be equal to the number of parallel paths K. By our ear-

lier observations, it suffices to identify the coloring subsets $\{(G_{i0}, G_{i1}, G_{i2}),\ \ \{F_i\} \mid 1 \leq i \leq K\}$ used by the protocol. These subsets are identified below:

$$G_i = [\{c_i\}, \{c_{i+1}\}, \{c_{i+2}\}], \tag{12}$$
$$F_i = \{c_{i+3}\}. \tag{13}$$

Verification that this coloring scheme meets all the constraints of Lemma 2.2 is straightforward. It follows that this orthogonal protocol achieves the cut-set bound . ■

The previous theorem handles the case when there are 4 or more backbone paths in the KPP network. When $K = 3$ it is not possible to identify efficient protocols that avoid back-flow. The lemma below reassures us that this does not prevent us from achieving the cut-set bound on the DMT.

Lemma 2.4: Consider a KPP network operating under an orthogonal protocol, which on neglecting the effect of back-flow, induces a block-diagonal channel matrix. IThen, the DMT of the protocol taking into account the effect of back-flow, is lower bounded by the DMT when neglecting the effect of back-flow .

Proof: The presence of back-flow creates entries in the strictly lower-triangular portion of the induced channel matrix. Since the DMT of a lower triangular matrix is lower bounded by the DMT of the corresponding diagonal matrix (by Theorem 1.3), we have that the system with back-flow will yield a DMT no worse than the one without back-flow. ■

We will now exploit Lemma 2.4 to construct a protocol achieving the cut-set bound for the KPP network with $K = 3$.

Theorem 2.5: For any KPP network with $K = 3$, there exists an explicit protocol achieving the cut-set bound on DMT.

Proof: For every parallel path, P_i, define

$$a_i = \begin{cases} 1, & n_i = 1 \text{ mod } 3 \\ 0, & n_i \neq 1 \text{ mod } 3 \end{cases}. \tag{14}$$

Without loss of generality we assume that the paths are ordered such that for the first l paths, $a_i = 1$ followed by the paths for which $a_i = 0$. We give a protocol for various possible values of l.

- *Case 1: ($l = 0, 1,$ or 3)*
  In this case, we will give a protocol that avoids back-flow, uses all paths equally, and achieves rate 1. By Lemma 2.2, this protocol will achieve the cut-set bound.
  We will specify the orthogonal protocol by specifying the activation sets $G_i = (G_{i0}, G_{i1}, G_{i2})$ and F_i for all i. We begin by setting $F_i = G_{i(n_i-1 \pmod 3)}$. The set of colors used is $C = \{c_0, c_1, c_2\}$.

  For $l = 0$,
  $$G_i = \begin{cases} (\{c_i\}, \{c_{i+2}\}, \{c_{i+1}\}), & n_i = 0 \text{ mod } 3 \\ (\{c_i\}, \{c_{i+1}\}, \{c_{i+2}\}), & n_i = 2 \text{ mod } 3 \end{cases}.$$

  For $l = 1$,
  $G_1 = (\{c_0\}, \{c_1\}, \{c_2\})],$
  $$G_2 = \begin{cases} (\{c_1\}, \{c_0\}, \{c_2\}), & n_2 = 0 \text{ mod } 3 \\ (\{c_1\}, \{c_2\}, \{c_0\}), & n_2 = 2 \text{ mod } 3 \end{cases},$$
  $$G_3 = \begin{cases} (\{c_2\}, \{c_0\}, \{c_1\}), & n_3 = 0 \text{ mod } 3 \\ (\{c_2\}, \{c_1\}, \{c_0\}), & n_3 = 2 \text{ mod } 3 \end{cases}.$$

  For $l = 3$,
  $G_i = (\{c_i\}, \{c_{i+1}\}, \{c_{i+2}\}).$

- *Case 2: ($l = 2$)*
  For $l = 2$, it turns out that there is no orthogonal protocol that satisfies all the conditions of Lemma 2.2. To handle this situation, we shall now come up with a protocol which satisfies conditions 1 and 2 of Lemma 2.2, but does not satisfy condition 3, i.e., the protocol will not avoid back-flow. Then, we utilize Lemma 2.4 to establish that the DMT for this protocol is equal to the cut-set bound. Consider the protocol having the following descriptor:
  $G_i = (\{c_i\}, \{c_{i+1}\}, \{c_{i+2}\})$ and $F_i = G_{i(n_i-1 \pmod 3)}$.

  After this assignment, we make the following modifications to A_{ij},

  $A_{3(n_3)} = \{c_2\},$
  $A_{3(n_3-1)} = \{c_0\}$, if $n_3 = 2 \pmod 3$.

  It can be checked that the protocol satisfies the first two conditions of Lemma 2.2. While condition 3 is not satisfied, because back-flow is present, by Lemma 2.4, the presence of back-flow does not worsen the DMT and therefore, the protocol achieves the cut-set bound on DMT.

■

C. The Case of Two Parallel Paths

We now proceed to the handle the last remaining case, namely the case when $K = 2$. As in the previous case, it turns out that it is impossible to have cut-set bound achieving orthogonal protocols that avoid back-flow. Furthermore, it is not even possible to construct a rate-1 orthogonal protocol with or without back-flow in general. We will first determine an upper bound to the rate of any orthogonal protocol for a KPP network with $K = 2$. This will then be followed by the presentation of an orthogonal protocol which achieves this upper bound.

Theorem 2.6: For any KPP network with $K = 2$, the maximum achievable rate for any orthogonal protocol is given by

$$R_{max} \leq \begin{cases} 1, & n_1 + n_2 = 0 \mod 2 \\ \frac{2n_2-1}{2n_2}, & n_1 + n_2 = 1 \mod 2 \end{cases}, \tag{15}$$

where $n_1 \leq n_2$.

Proof: Any given orthogonal protocol can be represented as a coloring of the edges satisfying the conditions in Prop. 1. For this case of $K = 2$, it will be found convenient to relabel the $n_1 + n_2$ edges in the network so as to form a cycle $l_1, l_2, ..., l_{n_1+n_2}$ of length $n_1 + n_2$. The specific relabeling is

given by

$$l_j = \left\{ \begin{array}{cc} e_{1j} \ , & j \leq n_1 \\ e_{2(n_2+n_1+1-j)} \ , & n_1 < j \leq n_1 + n_2 \end{array} \right. .$$

We associate edge $l_1, l_2, ..., l_{n_1+n_2}$ with color $D_1, D_2, ..., D_{n_1+n_2}$ so that

$$D_j = \left\{ \begin{array}{cc} A_{1j} \ , & j \leq n_1 \\ A_{2(n_2+n_1+1-j)} \ , & n_1 < j \leq n_1 + n_2 \end{array} \right. .$$

with a single constraint,

$$D_j \cap D_{(j+1) \pmod{(n_1+n_2)}} = \phi, \quad \forall j = 1, 2, ..., n_1 + n_2, \qquad (16)$$

that simultaneously meets the first three conditions laid out in Prop. 1.

Now suppose we have a coloring scheme with N colors. Then each color can be assigned to at most $\lfloor \frac{n_1+n_2}{2} \rfloor$ edges. This is because if more edges were assigned with the same color, then the half-duplex constraint would be violated. So by counting edge-color pairs in two different ways, we obtain the bound,

$$\begin{aligned} \sum_{i=1}^{2} \sum_{j=1}^{n_i} |A_{ij}| &\leq \left\lfloor \frac{n_1 + n_2}{2} \right\rfloor N, \\ \text{ie.}, n_1 m_1 + n_2 m_2 &\leq \left\lfloor \frac{n_1 + n_2}{2} \right\rfloor N, \qquad (17) \end{aligned}$$

where $m_i = |A_{ij}| \forall j \in [n_i]$. The half-duplex constraint implies that,

$$2m_1 \leq N, \qquad (18)$$
$$2m_2 \leq N. \qquad (19)$$

To find the maximum rate, we pose the maximization problem: maximize rate R = $(\frac{m_1}{N} + \frac{m_2}{N})$ subject to (17), (18), and (19). This can be easily solved to obtain,

$$\begin{aligned} \frac{m_1}{N} &= 0.5, \\ \frac{m_2}{N} &= \frac{1}{n_2} \left\lfloor \frac{n_1 + n_2}{2} \right\rfloor - \frac{n_1}{2n_2}. \end{aligned}$$

As a result, the maximum rate of the protocol is upper-bounded as

$$R_{max} \leq \left\{ \begin{array}{ll} 1, & n_1 + n_2 = 0 \pmod 2 \\ \frac{2n_2 - 1}{2n_2}, & n_1 + n_2 = 1 \pmod 2 \end{array} \right. , \qquad (20)$$

where $n_1 \leq n_2$.

■

Construction 2.7: This construction establishes an orthogonal protocol for $K = 2$ which achieves the maximum rate. By Prop. 1, it is sufficient to specify the coloring subsets A_{ij} $\forall i, j$ or equivalently, to specify the subsets D_j.

Case 1: $(n_1 + n_2) = 0 \pmod 2$

For this case we choose 2 as the cycle length of the protocol in our construction. Accordingly let the set of colors be $C = \{c_0, c_1\}$. Set

$$D_j = \left\{ \begin{array}{cc} \{c_0\} \ , & j = 1, 3, ..., n_1 + n_2 - 1 \\ \{c_1\} \ , & j = 2, 4, ..., n_1 + n_2 \end{array} \right. ,$$

which essentially corresponds to coloring the cycle formed by the network alternately with the two colors c_0 and c_1.

Case 2: $(n_1 + n_2) = 1 \pmod 2$

The coloring prescribed in Case 1 does not work here, since the cycle is of odd length. Therefore, we resort to a different coloring in this case.

We have the set of colors $C = \{c_1, c_2, ..., c_N\}$, where $N = 2n_2$. We will add colors to D_j using the following algorithm.

1) Step 1: $D_j \leftarrow \phi \ \forall j \in \{1, 2, ..., n_1 + n_2\}$.
2) Step 2: Now we will add colors to each of the set D_j using the following algorithm. In the algorithm, whenever we refer to D_j, with $j \notin \{0, 1, 2, \ldots, n_1+n_2\}$, we mean $D_j = D_{j \pmod{n_1+n_2}}$ and with $j = 0$, we mean $D_j = D_{n_1+n_2}$.

{
  $t \leftarrow 1$;
  for $k = 1$ to n_2 in steps of 1 :
  {
    for $i = 1$ to $n_1 + n_2 - 1$ in steps of 1 :
    {
      if i is odd, $D_{i-k+1} \leftarrow D_{i-k+1} \cup \{c_t\}$;
      if i is even, $D_{i-k+1} \leftarrow D_{i-k+1} \cup \{c_{t+1}\}$;
    }.
    $t \leftarrow t + 2$;
  }.
}.

Proposition 2: The orthogonal protocol shown in Construction 2.7 achieves the maximum rate given in Theorem 2.6.

Proof: For *Case 1*, it is clear that the algorithm achieves rate 1. For *Case 2*, the algorithm can be summarized as follows: During the i-th iteration, we fix our starting point as the i-th link in the longer path. We have two colors at the i-th iteration, C_{2i+1} and C_{2i+2}, which we will associate alternately with the edges in the circular loop starting from the $(i-1)$-th edge in the longer path P_2. This coloring respects the half-duplex, broadcast and interference constraints, because all the constraints reduce to a single one, i.e., adjacent edges in the network viewed as a cycle shall have distinct colors. After the n_2 iterations are over, we have mapped $2n_2$ colors to the network. The shorter path gets n_2 colors on each edge, whereas the longer path gets n_2-1 colors on each edge. Hence the rate of the resultant protocol will be equal to $\frac{2n_2-1}{2n_2}$. This is illustrated with an example, $(n_1, n_2) = (3, 4)$ in Fig. 9.

■

Theorem 2.8: For a KPP network with $K = 2$, if the two path lengths are equal modulo 2, then the DMT achieved by the orthogonal protocol of Construction 2.7 meets the cut-set bound, i.e., $d(r) = 2(1 - r)^+$.

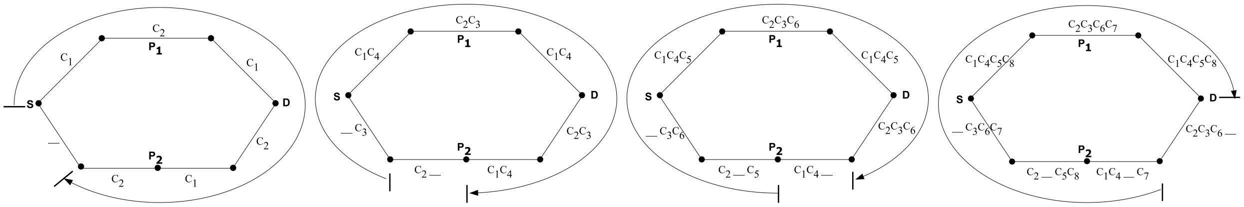
Proof: The proof follows from Lemma 2.2, Theorem 2.6 and Prop. 2. ■

Theorem 2.9: For a KPP network, there exists an orthogonal protocol achieving the cut-set bound if $K \geq 3$ or $K = 2$ and $n_1 = n_2 \pmod 2$.

Proof: Follows from Theorem 2.3, Theorem 2.5 and Theorem 2.8. ■

D. KPP Networks with Direct Link

Theorem 2.10: For KPP(D) networks, the cut-set bound on DMT is achievable whenever there is an orthogonal protocol



(a) Time slots 1 and 2 (b) Time slots 3 and 4 (c) Time slots 5 and 6 (d) Time slots 7 and 8

Fig. 9. Protocol Illustration: (n_1, n_2) = (3,4)

for the backbone KPP network satisfying the conditions of Lemma 2.2.

Corollary 2.11: For KPP(D) networks, the cut-set bound on DMT is achievable whenever $K \geq 4$.

Proof: (of Theorem 2.10)

The same protocol in the backbone KPP network yields an induced channel matrix, which is a diagonal matrix with the K path gains appearing cyclically along the diagonal. In the presence of a direct link, as is the case here, clearly it is the path gain $\mathbf{g}_d$ of the direct link that will appear along the diagonal of the induced channel matrix. The path gains of the backbone paths will appear in general below the main diagonal. However it is not hard to see that by suitably delaying symbols along each of the K paths, the path gains $\mathbf{g}_i, i = 1, 2, \ldots, K$ can be made to appear cyclically along a single sub-diagonal, say the D-th sub-diagonal.

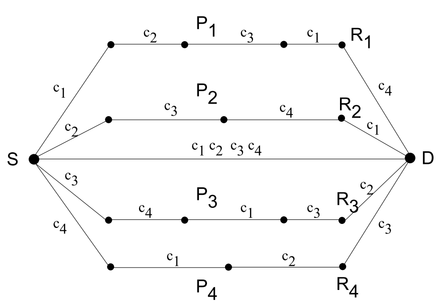


Fig. 10. KPP(D) network protocol

For instance, consider a KPP(D) network with 4 relaying paths (see Fig. 10). The figure describes an orthogonal protocol with the aid of a coloring using a set of colors $\{c_1, c_2, c_3, c_4\}$. The cycle length of the protocol is 4, and c_i represents the ith time-slot within a cycle. Unlike the case in KPP networks, there is no initial delay between the start of transmission at the source and the reception at the sink, due to the presence a direct link. However, the delays encountered by symbols in various backbone paths are potentially different. In this example, delay in paths P_1 and P_3 are 7 time-slots, whereas that in paths P_2 and P_4 are 3 time-slots. So if the fading coefficient along the path P_i is $\mathbf{g}_i$ and that along the direct link is $\mathbf{g}_d$, then the input-output relation between the first 10 transmitted and received symbols would be of the form:

$$
\begin{aligned}
\mathbf{y} &= \mathbf{H}\mathbf{x} + \mathbf{w}, \text{ where} \\
H &= \begin{bmatrix}
\mathbf{g}_d &&&&&&&&& \\
& \mathbf{g}_d &&&&&&&& \\
&& \mathbf{g}_d &&&&&&& \\
&&& \mathbf{g}_d &&&&&& \\
& \mathbf{g}_2 &&& \mathbf{g}_d &&&&& \\
&&&&& \mathbf{g}_d &&&& \\
&&& \mathbf{g}_4 &&& \mathbf{g}_d &&& \\
\mathbf{g}_1 &&&&&&& \mathbf{g}_d && \\
&&&&& \mathbf{g}_2 &&& \mathbf{g}_d & \\
&& \mathbf{g}_3 &&&&&&& \mathbf{g}_d
\end{bmatrix}.
\end{aligned}
$$

Here $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{w}$ represent the input, output and noise vectors respectively, each of length 10.

By adding a delay of 4 time-slots at R_2 and R_4, the effective delay encountered by symbols traveling via paths P_2 and P_4 can be made to equal 7 time-slots. The cycle length of the protocol being 4, a delay of 4 can be introduced by idling the respective nodes for one cycle of the protocol. This will result in a new channel matrix $\mathbf{H}'$ between the same input and output vectors $\mathbf{x}$ and $\mathbf{y}$, where

$$
\mathbf{H}' = \begin{bmatrix}
\mathbf{g}_d &&&&&&&&& \\
& \mathbf{g}_d &&&&&&&& \\
&& \mathbf{g}_d &&&&&&& \\
&&& \mathbf{g}_d &&&&&& \\
&&&& \mathbf{g}_d &&&&& \\
&&&&& \mathbf{g}_d &&&& \\
&&&&&& \mathbf{g}_d &&& \\
\mathbf{g}_1 &&&&&&& \mathbf{g}_d && \\
& \mathbf{g}_2 &&&&&&& \mathbf{g}_d & \\
&& \mathbf{g}_3 &&&&&&& \mathbf{g}_d
\end{bmatrix}.
$$

So the effective channel matrix due to an orthogonal protocol in a KPP(D) network will be of the form,

$$
\mathbf{H} = \begin{bmatrix}
\mathbf{g}_d &&&&& \\
& \mathbf{g}_d &&&& \\
&& \ddots &&& \\
\mathbf{g}_1 &&&&& \\
& \mathbf{g}_2 &&&& \\
&& \ddots && \ddots & \\
&&& \mathbf{g}_K && \mathbf{g}_d \\
&&&& \mathbf{g}_1 && \mathbf{g}_d
\end{bmatrix}.
$$

Next, consider the situation when the network is operated for a duration of $M = mK + D$ time-slots, for some positive

integer m. We now invoke Theorem 1.3 to the induced channel matrix $\mathbf{H}$ to arrive at a lower bound on the DMT as,

$$\begin{aligned} d_H(r) &\geq d_{H^{(0)}}(r) + d_{H^{(\ell)}}(r) \\ &\geq (1 - \frac{r}{M})^+ + K(1 - \frac{r}{M-D})^+. \\ \Rightarrow d(r) &= d_H(Mr) \\ &\geq (1-r)^+ + K(1 - \frac{M}{M-D}r)^+, \end{aligned}$$

which, as M tends to infinity, becomes

$$d(r) \geq (K+1)(1-r)^+. \tag{21}$$

For KPP(D) networks, the cut-set upper bound on DMT (Lemma 1.1) yields $d(r) \leq (K+1)(1-r)^+$. Combining this with the DMT lower bound (21), we get
$d(r) = (K+1)(1-r)^+$. ∎

Remark 9: Since a two-hop relay network possessing N relays with direct link and with relays isolated is a KPP(D) network with $K := N$, the DMT optimal strategy for these family of networks is given by Corollary 2.11. This turns out to be the same strategy as SAF protocol given in [10] for these networks.

III. Half-Duplex KPP(I) Networks

In this section, we move on to consider multi-hop networks with interference. We consider KPP(I) networks with single antenna nodes operating under the half-duplex constraint. We will show that even here, the cut-set bound can be achieved using AF protocols. The cut-set bound (Lemma 1.1) gives the same DMT upper-bound $d(r) \leq K(1-r)^+$, as in the case of KPP networks. In this section, we will demonstrate that this DMT is in fact achievable.

In the case of a KPP(I) networks, we note that any protocol for the backbone KPP network automatically induces a protocol on the KPP(I) network. Although a protocol is orthogonal with respect to the backbone KPP network, it will most likely result in a protocol on the KPP(I) network that is not orthogonal because in the presence of interfering links, interference avoidance is no longer guaranteed. The aim here is to come up with a protocol for the backbone KPP network that induces a derived protocol on the KPP(I) network such that it will result in a DMT-achieving channel matrix for the KPP(I) network. As explained in the case of KPP networks, without loss of optimality, we can neglect the initial delay here also while considering the effective channel matrix induced by the protocol between the source and the sink. We begin by introducing the notion of causal interference.

A. Causal Interference

Definition 9: Let $\mathbf{H}$ denote the channel matrix induced by an AF protocol $\wp$ employed in a KPP(I) network $\mathcal{N}$. Let $\mathbf{H}^{(0)}$ be the diagonal part of $\mathbf{H}$. Let $\mathbf{H}_{bb}$ denote the channel matrix induced by $\wp$ when it is employed on the backbone of $\mathcal{N}$. Then $\wp$ is said to be a causal protocol (i.e., a protocol with causal interference) if the interference admitted by the protocol is causal in nature, i.e.,

1) $\mathbf{H}$ is lower triangular and
2) $\mathbf{H}_{bb} = \mathbf{H}^{(0)}$.

Remark 10: Note that condition 2) in the definition above is satisfied only if the protocol $\wp$ is derived from a protocol that is orthogonal with respect to the backbone KPP network. For this reason, all the protocols $\wp$ for KPP(I) networks encountered in this section will be derived from orthogonal protocols for the backbone KPP network. Thus throughout the remainder of this section, whenever the word protocol appears in the context of KPP(I) networks, it should be interpreted to mean a protocol derived from an orthogonal protocol for the backbone KPP network.

Lemma 3.1: Consider a KPP(I) network operating under a causal protocol $\wp$. Let the induced channel matrices on the KPP(I) and backbone KPP networks be given by $\mathbf{H}$ and $\mathbf{H}_{bb}$ respectively. Then the DMT of $d_H(r) \geq d_{H_{bb}}(r)$. Furthermore, if the protocol achieves the cut-set bound on the backbone KPP network, i.e., $d_{H_{bb}}(r) = K(1-r)^+$, then it also achieves the cut-set bound on the KPP(I) network, i.e., $d_H(r) = K(1-r)^+$.

Proof: Since the interference caused by the protocol is causal, $\mathbf{H}$ is lower triangular. We also have that $\mathbf{H}^{(0)} = \mathbf{H}_{bb}$ is the diagonal part of $\mathbf{H}$. The lemma now follows because by Theorem 1.3, the DMT of a lower triangular matrix is lower bounded by the DMT of the corresponding diagonal matrix, i.e., $d_H(r) \geq d_{H_{bb}}(r)$. The rest follows since the same cut-set bound also applies to the KPP(I) network. ∎

We now give a sufficient condition for a protocol to be causal.

Proposition 3: Consider a KPP(I) network under a protocol $\wp$. Then $\wp$ is causal if the unique shortest delay experienced by every transmitted symbol is through a backbone path. ∎

Proof: The proof is straightforward. ∎

Proposition 4: Consider a KPP(I) network under a protocol $\wp$ derived from a protocol that achieves the cut-set bound on the KPP network. If the unique shortest delay experienced by every transmitted symbol is through a backbone path, then $\wp$ achieves the optimal DMT of $d(r) = K(1-r)^+$.

Proof: The proof is clear by combining Prop. 3 and Lemma 3.1. ∎

B. Optimal DMT for Regular Networks

For the class of regular networks, the sufficient condition given in Prop. 4 is easily satisfied, leading to the following result.

Theorem 3.2: The optimal DMT $d(r) = K(1-r)^+$ of (K, L) regular networks is achievable.

Proof: Consider a (K, L) regular network. This network can be regarded as a KPP(I) network in which each backbone path has an equal number of edges. We operate the network using a protocol that is derived from an orthogonal protocol specified for the backbone KPP network. The orthogonal protocol for the backbone KPP is specified by giving the coloring sets. The cycle length of the protocol is K, and hence we use the set of colors $C = \{c_0, c_1, \ldots, c_{K-1}\}$, and assume $\forall \ell > K, c_\ell = c_{\ell \pmod K}$. The coloring set for edge e_{ij} is given by,

$$A_{ij} = \{c_{i+(j-1)}\}, \quad 1 \leq i \leq K, \; 1 \leq j \leq L+1. \tag{22}$$

An example of this coloring scheme for the network in Fig. 6 is given in Fig. 11.

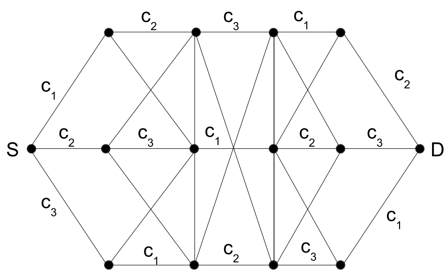


Fig. 11. Regular network with the protocol

It can be verified that the shortest delay encountered by every symbol is through a backbone path. It can also be checked that the protocol satisfies conditions of Lemma 2.2 and thus achieves the cut-set bound on the backbone KPP network. Thus the conditions of Prop. 4 are satisfied and hence proposed protocol achieves the cut-set bound on the (K, L) regular network. ■

Corollary 3.3: For the two-hop relay network without direct link, the optimal DMT is achieved irrespective of the presence or absence of links between relays.

Proof: The two-hop relay network without the direct link is a (K,1) regular network, where K denotes the number of relays in the network. This holds irrespective of the presence of links between relays, since they only contribute to intra-layer links. Thus Theorem 3.2 implies this corollary. ■

Remark 11: The result in Corollary 3.3 was also proven in a parallel work [19]. The protocols used in this paper and in [19] are essentially the same as the slotted-amplify-and-forward (SAF) protocol [10], except that the protocol is applied to a network that does not have a direct link. However, the proof techniques used here and in [19] are very different.

C. Causal Interference for KPP(I) Networks with $K = 3$

In this subsection, we construct a protocol satisfying the conditions of Prop. 4 and thereby, the protocol will be optimal. The protocol will turn out to require the introduction of suitable delays at intermediate nodes in the network. We begin by considering KPP(I) networks with $K = 3$. Subsequently, we will use the theory developed here for $K = 3$ to handle the case $K > 3$.

Lemma 3.4: For any KPP(I) networks with $K = 3$, there exists a causal protocol that achieves the cut-set bound when used on the back-bone KPP network.

Proof: The proof is deferred to Appendix I. ■

Theorem 3.5: For any KPP(I) network with with $K = 3$, there exists an explicit protocol that achieves the cut-set bound on the DMT.

Proof: In Lemma 3.4, we showed how to construct an orthogonal protocol that admits causal interference. This lemma along with Lemma 3.1 proves the theorem. ■

D. Optimal DMT of KPP(I) networks

In the last section, we have shown how to design protocols for KPP(I) networks which admit causal interference. This leads to the following results on the optimal DMT of KPP(I) networks with $K \geq 3$.

Theorem 3.6: Consider any KPP(I) network with $K \geq 3$. The cut-set bound on the DMT $d(r) = K(1-r)^+$ is achievable.

Proof: For $K = 3$, it follows from Theorem 3.5.

Now, we will consider the case when $K > 3$. First we use the procedure illustrated in Lemma 1.1 in order to remove all non-contiguous k-switches for $2 \leq k \leq K$. This will lead us to a modified KPP(I) network which does not contain any non-contiguous k-switches. Consider a 3PP sub-network of this KPP(I) network. From Lemma 3.4, it follows that we can design a causal protocol for such a network which results in an induced channel matrix with three product coefficients corresponding to the three back-bone paths along the diagonal. There are now $\binom{K}{3}$ possible 3PP subnetworks of the modified KPP(I) network. If each of these subnetworks is activated in succession, it would yield a lower triangular matrix with all the K product coefficient g_i repeated $\binom{K-1}{2}$ times along the diagonal. By Theorem 1.3, the DMT of this matrix is no worse than that of the diagonal matrix alone. The diagonal matrix has a DMT equal to $K(1-r)^+$ after rate normalization. Therefore a DMT of $d(r) \geq K(1-r)^+$ can be obtained. However, since $d(r) \leq K(1-r)^+$ by the cut-set bound, we have $d(r) = K(1-r)^+$. ■

IV. Half-Duplex Layered Networks

In this section, we consider half-duplex layered networks with single-antenna nodes. As in the case of KPP(I) networks, an achievable DMT for layered networks with full-duplex relays comes as an immediate consequence of Theorem 1.7. It is given in the below corollary.

Corollary 4.1: For the full-duplex layered networks, a linear DMT between the maximum diversity and maximum multiplexing gain can be achieved.

Half-duplex layered networks are typically treated by first considering full-duplex operation and then activating alternate layers in order to satisfy the half-duplex constraint [9], [21]. However, this leads to a rate loss of a factor of two. We will demonstrate in this section that there exists schedules for which this rate loss is not incurred.

We will prove that the same result holds good for fully-connected (fc) layered networks even when the relays are of half-duplex nature. First, we will establish a sufficient condition for a layered network, such that a linear DMT of $d_{\max}(1-r)^+$ is achievable. Also, we will prove that $d_{\max}(1-r)^+$ is always achievable for the fc layered networks.

A. Linear DMT in Layered Networks

Similar to the case of KPP networks, the basic idea is to activate $d_{\max}$ paths from the source to the sink using AF protocol. In the case of KPP networks, clearly we have $d_{\max} = K$ node-disjoint paths, and we could activate them without interference between one another. In the case of layered networks, we begin with identifying node-disjoint paths from all possible paths from the source to the sink. A path from source to sink in a layered network is said to be

forward-directed if all the edges in the path are directed from one layer to the next layer towards the sink (i.e., no edge in the path goes from one layer to the previous layer and there is no edge which starts and ends in the same layer). Identification of node-disjoint forward-directed paths will allow us to schedule the edges in the network in a similar fashion as how the parallel paths in KPP network were scheduled. In the following lemma, we propose a technique to identify node-disjoint paths in a layered networks.

Definition 10: Given a set of forward-directed paths P in a layered network, the bipartite graph corresponding to P is defined as follows:

- Construct a bipartite graph with vertices corresponding to paths in P on both sides.
- Connect a vertex associated with path P_i on the left to a vertex associated with P_j on the right if the two paths are node disjoint.

Lemma 4.2: Consider a set of paths $P = \{P_i, i = 1, 2, \ldots, N\}$ in a given layered network. Let the product of the fading coefficient on the i-th path P_i be $\mathbf{g}_i$. Construct the bipartite graph corresponding to P. If there exists a complete matching in this bipartite graph, then these paths can be activated in such a way that the DMT of this protocol is greater than or equal to the DMT of a parallel channel with fading coefficients $\mathbf{g}_i, i = 1, 2, ..., N$ with the rate reduced by a factor of N, i.e., $d(r) \geq d_{H_d}(Nr)$, where $H_d = \text{diag}(\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N)$.

Proof: Suppose there is a complete matching π on the graph constructed as above. The complete matching specifies for every path on the left P_i, a partner on the right P_{π_i}. The length of each path and therefore the delay is equal to $D := L + 1$ since the network is layered with L relaying layers. Consider operating the network under the following protocol.

Step 1: Set $i = 1$. Consider the path P_i along with its partner path P_{π_i} as a two parallel path KPP(I) network with backbone paths of equal length. This sub-network is in fact a $(2, L)$ regular network with L being the number of layers in the original network. The sub-network is activated using the protocol specified in the proof of Theorem 3.2 for $T >> 2$ cycles, and since each protocol cycle is of duration 2 time slots, the total length of activation is $2T$ time slots. Since the protocol is causal, it will induce a lower triangular channel matrix between the input and the output with channel gains $\mathbf{g}_i$ and $\mathbf{g}_{\pi_i}$ alternating along the diagonal.

Step 2: Repeat *Step 1* for $i = 2, 3, \ldots, N$.

The induced channel matrix $\mathbf{H}$ will be a lower triangular $2NT \times 2NT$ matrix with each of $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N$ appearing for $2T$ time-slots.

Now the DMT of the protocol $d(r)$ can be related to the DMT of the diagonal part $\mathbf{H}^{(0)}$ as $d(r) = d_H(2NTr) \geq d_{H^{(0)}}(2NTr)$ by applying Theorem 1.3. If $\mathbf{H}_d$ is a $N \times N$ diagonal matrix comprising of $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N$ along the diagonal, then $d_{H^{(0)}}(r) = d_{H_d}(\frac{r}{2T})$ by applying Lemma 1.5. This yields $d(r) \geq d_{H_d}(Nr)$. ∎

We utilize the above lower bound to obtain a sufficient condition that guarantees that a linear DMT between the maximum diversity and multiplexing gain is achieved on a general layered network in the following lemma.

Lemma 4.3: For a general layered network, a DMT of $d(r) \geq M(1 - r)^+$ is achievable whenever the network has M forward-directed edge-disjoint paths from the source to the sink and the bipartite graph corresponding to the set of edge-disjoint paths e_i, $i = 1, 2, ..., M$ has a complete matching.

Remark 12: Since an arbitrary network without a direct link can be represented as a general layered network by Remark 2, this lemma applies to arbitrary networks. If in addition to the requirements in the lemma above, the network has the number of forward-directed edge-disjoint paths equal to the min-cut, then a linear DMT of $d_{\max}(1 - r)^+$ is achievable.

Proof: (of Lemma 4.3) By using Lemma 4.2 we will be able to get a DMT of $d(r) \geq d_{H_d}(Mr)$. But since the paths are edge-disjoint, the fading coefficients are independent, and we get $d_{H_d}(r) = (M - r)^+$ by applying Lemma 1.5. Therefore, we get, $d(r) \geq M(1 - r)^+$. ∎

Since the min-cut M is equal to the maximum diversity of the network by Theorem 1.6, a DMT of $M(1 - r)^+$ signifies a DMT that is linear between the maximum diversity and the maximum multiplexing gain 1.

B. Fully-Connected Layered Networks

While we have established a sufficient condition for achieving a linear DMT between maximum diversity and multiplexing gain for an arbitrary layered network, even for many fc layered networks, this sufficient condition is not satisfied. For example, a fc layered network with $(3, 2, 3)$ nodes does not satisfy the sufficient condition in Lemma 4.3. However, as we shall see in the sequel, this can be remedied and a linear DMT is obtainable for arbitrary fc layered networks with number of nodes in any relaying layer being at least 2. A supporting lemma is needed before we proceed to prove this main result.

Consider an fc layered network with L relaying layers. Let $\mathbf{h}_j^{(i)}$ be the i-th fading coefficient in the j-th hop. Let there be R_i relay nodes in the i-th layer. Then the number of forward-directed paths N is equal to

$$N = \prod_{i=1}^{L} R_i. \tag{23}$$

Lemma 4.4: Consider an fc layered network with L relaying layers. Let there be R_i relay nodes in the i-th layer. Let $P = \{P_1, .., P_N\}$ be the set of all forward directed paths in the network. Then the bipartite graph corresponding to P has a complete matching.

Proof: We will prove this by producing an explicit complete matching on the bipartite graph. Let the layered network have L layers. Let us fix an (arbitrary) ordering on the relays in each hop. Let the relays in the j-th hop be indexed $0, 1, ..., R_j - 1$.

There is a one-to-one correspondence between the set of all forward-directed paths and the L tuples $(b_1, ..., b_L)$, where b_j denotes the index of the relay in the j-th layer visited by that path. For any given path P associated to the L tuple $(b_1, ..., b_L)$, consider a path P' associated to the tuple $(c_1, ..., c_L)$ where $c_i = b_i + 1 \pmod{R_i}$. Clearly these two paths are node-disjoint, because $R_i \geq 2, \ \forall i$ by definition of fc layered network. The collection of all such pairings constitutes

a complete matching on the set of all forward-directed paths. ■

Consider an fc layered network with N forward-directed paths and L relaying layers. Let $\mathbf{g}_i$ denote the product fading coefficient corresponding to the forward-directed path P_i for $i = 1, 2, \ldots, N$. Now, $\{\mathbf{g}_i\}$ are mutually correlated because each $\mathbf{g}_i$ is the product of Rayleigh fading coefficients of links in that path, and a link may belong to multiple paths. Next we compute the DMT of a parallel channel with N sub-channels in which ith sub-channel corresponds to the ith forward-directed path in the fc layered network. Clearly, the coefficient of the ith sub-channel is $\mathbf{g}_i$, which is the product of $L+1$ Rayleigh coefficients, one each chosen from $L+1$ different sets of Rayleigh coefficients. Here each of the $L+1$ sets corresponds to the set of fading coefficients of all links connecting two adjacent layers of relays.

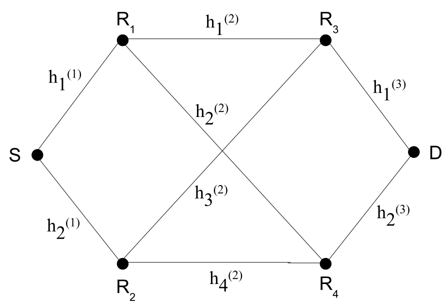


Fig. 12. Three hop layered network

For instance, consider an fc layered network with two layers of relays. [5] (see Fig. 12) There are four forward-directed paths between the source and the sink. The product coefficients corresponding to these four paths are

$$\mathbf{g}_1 = \mathbf{h}_1^{(1)}\mathbf{h}_1^{(2)}\mathbf{h}_1^{(3)} \quad (24)$$

$$\mathbf{g}_2 = \mathbf{h}_1^{(1)}\mathbf{h}_2^{(2)}\mathbf{h}_2^{(3)} \quad (25)$$

$$\mathbf{g}_3 = \mathbf{h}_2^{(1)}\mathbf{h}_3^{(2)}\mathbf{h}_1^{(3)} \quad (26)$$

$$\mathbf{g}_4 = \mathbf{h}_2^{(1)}\mathbf{h}_4^{(2)}\mathbf{h}_2^{(3)}. \quad (27)$$

Our interest is to compute the DMT of the parallel channel with 4 sub-channels with coefficients $\mathbf{g}_1, \ldots, \mathbf{g}_4$. The DMT can not be computed using Lemma 1.4 since the individual links are not independent. Intuitively, the diversity of this channel is at most two, since if $\mathbf{h}_1^{(1)}$ and $\mathbf{h}_2^{(1)}$ are both in deep fade, then the overall channel is likely in deep fade. We present the lemma below, which computes the DMT of general parallel channels with product coefficients having a certain structure. As a result of this lemma, it will follow that the example channel has a diversity of two.

Definition 11: A subset S of a product set $\mathcal{H} \subset \{H^{(1)} \times H^{(2)} \times \cdots \times H^{(L+1)}\}$ is said to be $(\nu_1, \nu_2, ..., \nu_{L+1})$-balanced if every element in $H^{(i)}$ appears precisely ν_i times as the ith coordinate of an element in S. Clearly we have $|S| = \nu_i |H^{(i)}|$.

Lemma 4.5: Let $H^{(i)} = \{\mathbf{h}_1^{(i)}, \mathbf{h}_2^{(i)}, \ldots, \mathbf{h}_{M_i}^{(i)}\}, i = 1, 2, ..., L+1$ be sets of i.i.d. Rayleigh fading coefficients. Now let $\mathcal{H} \subset \{H^{(1)} \times H^{(2)} \times \cdots H^{(L+1)}\}$ be a $(N_1, N_2, ..., N_{L+1})$-balanced set and let $N = |\mathcal{H}|$. Let $N_{\max} = \max_{i=1,2,\ldots,N} N_i$ and $M_{\min} = \min_{i=1,2,\ldots,L+1} M_i$. Let $\psi : \mathcal{H} \to G$ be the product map such that $\psi((a_1, a_2, ..., a_{L+1})) = \prod_{j=1}^{L+1} a_i$. Now let $\psi(\mathcal{H}) = \{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N\}$, where $\mathbf{g}_i = \prod_{k=1}^{L+1} \mathbf{h}_{e(i,k)}^{(k)}$, with $e(i,k)$ being a map from $[N] \to [M_k]$ for a fixed $k \in [L+1]$.

Let $\mathbf{H}$ be a $N \times N$ diagonal matrix with the diagonal elements given by $\mathbf{g}_i$. The DMT of the parallel channel $\mathbf{H}$ is a linear DMT between a diversity of $\frac{N}{N_{\max}}$ and a multiplexing gain of N,

$$d(r) = \frac{(N-r)^+}{N_{\max}}. \quad (28)$$

Proof: The proof is deferred to Appendix II. ■

Having thus established the DMT of a parallel channel with product fading coefficients, we now proceed to utilize this lemma to compute a lower bound on the DMT of fc layered networks.

Theorem 4.6: For a fully-connected layered network, a linear DMT between the maximum diversity and maximum multiplexing gain of 1 is achievable.

Proof: Consider an fc layered network with L layers. Let there be R_i antennas in the i-th layer for $i = 0, 1, ..., L+1$. We consider the source as layer 0, and sink as the $L+1$th layer so that $R_0 = R_{L+1} = 1$. Let $M_i := R_{i-1}R_i, i = 1, 2, ..., L+1$ be the number of fading coefficients between the $(i-1)$-th and ith layer of relays.

Let $H^{(i)} := \{\mathbf{h}_j^{(i)}, j = 1, 2, .., M_i\}$ be fading coefficients of links connecting nodes in the $(i-1)$th layer to those in the ith layer, $i = 1, 2, ..., L+1$. Let N be the total number of forward-directed paths from source to sink, and $P_i, i \in [N]$ be the various forward-directed paths. Let P denote the set of all these forward-directed paths. Then $|P| = N = \prod_{i=1}^{L} R_i$. Let $\mathbf{g}_i$ be the product fading coefficient on path P_i. Let $M_{\min} = \min_{i=1}^{L+1} M_i$ and $N_{\max} = \max_{i=1}^{L+1} N_i$. Note that $M_{\min}$ corresponds to the value of the min-cut.

By Lemma 4.4, the bipartite graph corresponding to P has a complete matching. The set of paths P satisfies the criterion of Lemma 4.2 and therefore, we can obtain a DMT of

$$d(r) \geq d_{H_d}(Nr) \quad (29)$$

where $\mathbf{H}_d$ is an $N \times N$ diagonal matrix whose ith diagonal entry is $\mathbf{g}_i$. Now, we need to compute $d_{H_d}(r)$. We associate with every path P_i an $L+1$ tuple of fading coefficients $\Theta_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{i(L+1)})$, where $\theta_{ik} \in H^{(k)}$ is the fading coefficient of the kth link in path P_i. Now $\mathbf{g}_i$ is related to Θ_i as $\mathbf{g}_i = \prod_{k=1}^{L+1} \theta_{ik}$. Note that the collection of fading coefficient tuples Θ_i corresponding to all the paths $P_i \in P$ is a $(N_1, N_2, \ldots, N_{L+1})$-balanced subset of the cartesian product set $\{H^{(1)} \times H^{(2)} \times \cdots \times H^{(L+1)}\}$, where

$$N_\ell = \frac{\prod_{j=0}^{L+1} R_j}{R_{\ell-1}R_\ell} \quad (30)$$

represents the number of tuples Θ_i in which h_j^i appears as a component for any $j \in [M_\ell]$.

We can now see that the parallel channel matrix $\mathbf{H}_d$ satisfies the conditions of Lemma 1.4, which can be applied to obtain

[5] This network was also used in [20] for illustration.

the DMT of $\mathbf{H}_d$ as

$$d_{H_d}(r) = \frac{(N-r)^+}{N_{\max}}. \quad (31)$$

Substituting this back in (29), we obtain a lower bound to the DMT of the protocol as

$$\begin{aligned} d(r) &\geq d_{H_d}(Nr) && (32)\\ &= N\frac{(1-r)^+}{N_{\max}} && (33)\\ \Rightarrow d(r) &\geq M_{\min}(1-r)^+. && (34)\end{aligned}$$

By Theorem 1.6, we have that the maximum diversity is equal to the min-cut $d_{\max} = M_{\min}$ in any network. Therefore a DMT of

$$d(r) \geq d_{\max}(1-r)^+ \quad (35)$$

is achievable. ■

Therefore for the two-layer fc layered network example considered in Fig. 12, a DMT of $2(1-r)^+$ is achievable using the strategy proposed in the theorem. However this DMT is also the upper bound using the cut-set bound. Therefore the DMT for this example network is identically equal to $2(1-r)^+$. This is not particular to this example. For fc layered networks with $L < 4$, the min-cut is either at the source side or at the sink side, and hence we have the following corollary:

Corollary 4.7: For an fc layered network with $L < 4$, the cut-set bound on DMT is achievable.

Proof: Let n_i be the number of relays in layer i. Consider a layered network with $L = 1$, i.e., there is only one layer. The DMT upper bound is $n_1(1-r)^+$ from the cut-set bound, which is achieved using the proposed strategy. For $L = 2$, the cut-set bound on DMT is $\min\{n_1, n_2\}$ $(1-r)^+$, which is achieved. For $L = 3$, it can be seen that $d_{\max} = \min\{n_1, n_2\}$ and that the DMT upper bound is $\min\{n_1, n_2\}$ $(1-r)^+$, which is indeed achieved. ■

V. Networks with Multiple Antenna Nodes

In the previous sections, we considered networks with all nodes having single antennas. In this section, we consider the general case of networks with multiple antenna nodes. Specifically, we consider KPP, KPP(I) and fc layered networks under both full-duplex and half-duplex constraints. As it is difficult to characterize the DMT completely, we present lower bounds, i.e., an achievable DMT region.

A. Full-Duplex, Fully-Connected Layered Networks

We consider fc layered networks with multiple antennas at the source and the sink. In the case of networks with multiple antennas at relays, we will replace every multiple-antena relay with an equivalent number of single-antenna relays in the same layer, and proceed to analyze the resultant network. Clearly, any protocol on the resultant network has a derived protocol in the original network and clearly the DMT of the resultant network serves as a lower bound to the DMT of the original network.

Definition 12: A ss-ss layered network with multiple antennas at the source and the sink is referred to as an

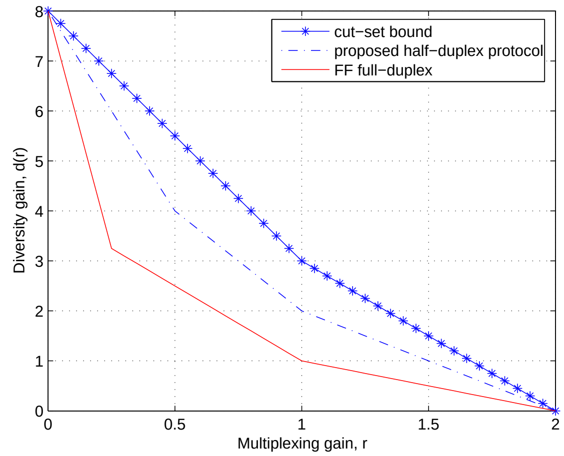


Fig. 13. Comparison of various protocols for $(2, 4, 2)$ network

$(n_0, n_1, \ldots, n_L, n_{L+1})$ network if the network has L layers, with the source having n_0 antennas, the sink having n_{L+1} antennas, and the i-th layer of relays having n_i single-antenna nodes.

In [9], parallel AF and flip-and-forward (FF) protocols have been proposed for the $(n_0, n_1, \ldots, n_{L+1})$ network with full-duplex operation and directed antennas. The parallel AF protocol aims to achieve the full diversity for the network, whereas FF achieves the extreme points of maximum multiplexing gain and the maximum diversity gain. In [9], it has been shown that the FF protocol achieves a better DMT than does an AF protocol. However, the DMT curves of both these protocols lie some distance away from the cut-set DMT bound. In both parallel AF and FF protocols, the key idea is to partition the relay nodes in each layer into subsets of nodes called super nodes. We propose a protocol with achievable DMT (i.e, with a lower bound to DMT) that is better than that of existing protocols for a full-duplex, $(n_0, n_1, \ldots, n_{L+1})$ network.

A "partitioning" $\mathfrak{p}$ of a layered network corresponds to dividing the nodes in every layer into super-nodes. We refer to the number of nodes within a super-node as the size of the super-node. Partitioning could potentially include partitioning of the source by which we mean dividing the transmit antennas of the source into super-nodes. It is similarly possible to partition the sink. Connecting any two super-nodes of size a and b respectively, are ab edges, which we will regard as a single super-edge. We will term the resultant network comprised of super-nodes connected through super-edges as the super-network. The super-network inherits from the original network the property of being fully-connected. For a given partitioning $\mathfrak{p}$, let $S_i^{(\mathfrak{p})}$ be the number of super-nodes in layer i. The number of super-edge-disjoint paths in the super-network is equal to the min-cut of the super-network given by,

$$N^{(\mathfrak{p})} := \min_{\{i=0,1,2,\ldots,L\}} S_i^{(\mathfrak{p})} S_{i+1}^{(\mathfrak{p})}.$$

We propose a protocol which uses different partitionings depending upon the multiplexing gain r (we will refer to

r as the rate by abuse of notation). [6] The basic intuition is that, at lower rates, we can exploit the diversity of the network by using a partitioning that results in the creation of a large number of super-edge-disjoint paths. At higher rates, we utilize a partitioning which supports enough degrees of freedom (which is equal to the minimum of the number of antennas in any super-node of an edge-disjoint path).

Let $\mathfrak{P}$ denote all possible partitionings. There can be two different partitionings in which every layer has the same number of super-nodes with corresponding sizes. However, we need to choose only one of them, as the performance of the protocol depends only on the size of super-nodes. Fix a partitioning $\mathfrak{p} \in \mathfrak{P}$ on the layered network. We operate the network using the following protocol: Activate all the $N^{(\mathfrak{p})}$ super-edge-disjoint paths successively so that each path is activated for T time instants. During the activation of ith path, we will get an induced channel matrix $\mathbf{H}$ that is block lower-triangular with $\mathbf{H}_i^{(\mathfrak{p})}$, the product matrix for the i-th path appearing as the ith entry along the diagonal of $\mathbf{H}$. The induced matrix has entries in the lower triangular part of the matrix due to the presence of back-flow.

Since $\mathbf{H}$ is blt, by Theorem 1.3, we can lower bound the DMT of this matrix by the DMT of $\mathbf{H}^{(0)}$, the block diagonal matrix extracted from $\mathbf{H}$. Let $d_i^{(\mathfrak{p})}(r)$ be the DMT of $\mathbf{H}_i^{(\mathfrak{p})}$, which can be computed using the techniques for computing the DMT of product Rayleigh matrices given in [9]. Now the DMT $d_{H^{(0)}}(r)$ can be computed using the parallel channel DMT in Lemma 1.4 to yield the DMT of the protocol as,

$$d(r) \geq \max_{\{\mathfrak{p}\in\mathfrak{P}\}} \inf_{\left\{\begin{array}{c}(r_1, r_2, \cdots, r_{N^{(\mathfrak{p})}}): \\ \sum_{i=1}^{N^{(\mathfrak{p})}} r_i = N^{(\mathfrak{p})} r\end{array}\right\}} \sum_{i=1}^{N^{(\mathfrak{p})}} d_i^{(\mathfrak{p})}(r_i). \tag{36}$$

Since the optimization is over the set of all possible partitions, it might be difficult to compute the DMT in general. So we consider a restricted case when the source and sink are not partitioned, and all the relay layers are partitioned into the same number of super-nodes, S. Under this assumption, we have that $1 \leq S \leq n_{\min}$, where $n_{\min} = \min\{n_i\}$ is the minimum number of antennas in any layer. Each super-node in the ith layer contains n_i^S relays, where $n_i^S := \left\lfloor \frac{n_i}{S} \right\rfloor, i = 1, 2, ..., L$. The remaining $n_i \pmod S$ relays in layer i are requested to be silent. This is done for simplicity of computing the DMT. Let $d_{(n_0,n_1,...,n_{L+1})}(r)$ denote the DMT of a product of independent Rayleigh matrices of size $n_0 \times n_1, n_1 \times n_2, ..., n_L \times n_{L+1}$, which can be computed using the techniques given in [9]. The DMT lower bound in (36) now simplifies to

$$d(r) \geq \max_{S\in\{1,2,...,n_{\min}\}} S\ d_{(n_0,n_1^{(S)},...,n_L^{(S)},n_{L+1})}(r). \tag{37}$$

The strategy of Corollary 4.1 can be used to obtain a DMT of $d_{\max}(1-r)^+$ for any full-duplex layered network even in the presence of multiple antennas at source and sink. By combining this strategy with the aforementioned strategy and choosing the one with the better DMT based on r, we get a DMT of

$$\begin{aligned} d(r) \ \geq\ \max\{&d_{\max}(1-r)^+, \\ &\max_{S\in\{1,2,...,n_{\min}\}} S\ d_{(n_0,n_1^{(S)},...,n_L^{(S)},n_{L+1})}(r)\}. \end{aligned} \tag{38}$$

The proposed protocol is essentially the same as [9] except for the following differences:

- The presentation here is not restricted to directed graphs since we are able to handle back-flow that might arise in an undirected graph by using Theorem 1.3 to show that back-flow does not impair the DMT.
- The presentation here is not restricted to partitions of fixed size since evaluation of the DMT in the case of arbitrary size partitions is made possible by the use of Lemma 1.4, which computes the DMT of the parallel-channel.
- Also, since we permit the size of the partition to vary with the rate, the additional flexibility can be used to improve upon the DMT attained by the FF protocol. While the dependence of network operation upon the rate could increase implementation complexity, one could adopt an intermediate strategy in which there are a small number of modes of operation, for example, two modes reserved respectively for low and high-rates.
- Finally, as will be shown in the sequel, the above results can be extended to half-duplex networks when all the relay layers are partitioned into the same number of super-nodes.

Example 1 : Consider a $(2, 4, 2)$ multi-antenna layered network. The achievable DMT curve using the FF protocol, the proposed protocol and the cut-set bound are plotted in the Figure 13. For rates $r \leq 0.5$, the strategy for full-duplex layered network given in Corollary 4.1 without any node-partitioning performs the best. For rates $r \geq 0.5$, partitioning the middle layer into two super-nodes with each containing two nodes performs better. A combination of these two strategies gives a superior DMT performance to the existing FF protocol.

B. Half-Duplex Fully-Connected Layered Networks

We consider multi-antenna layered networks with the additional constraint of half-duplex relay nodes. We prove that the methods provided above for full-duplex networks can be generalized for the half-duplex network with bidirectional links.

Consider the partitioning method stated for full-duplex layered networks, with $S_i = S, \forall i = 1, 2, ..., L$, i.e., the relaying layers are partitioned into equal number of super-nodes. Let the source and sink be un-partitioned. When the relay layer i is partitioned into S_i partitions, each super-node contains $n_i^S := \left\lfloor \frac{n_i}{S_i} \right\rfloor$ relays. If it contains more, the remaining relays are requested to be silent, as in the full-duplex case.

The following observations are in place: Once we replace the nodes corresponding to the same partition by a super-node, this virtual network forms a regular network. This is because each relaying layer has the same number of partitions

[6] The idea of varying the protocol parameters depending on the multiplexing gain r was used in [4] for the NSDF protocol.

and therefore the same number of super-nodes. The resultant network being regular, we use the protocol that is given in Theorem 3.2. Since the paths are of equal length, the interference is causal, making the induced channel matrix lower triangular. This has better DMT than the corresponding diagonal matrix by Theorem 1.3. This yields the same lower bound on DMT as in the full-duplex case. Thus the DMT of the protocol with any given partitioning in the half-duplex case is no worse than that with full-duplex protocol under the same partitioning. So we get,

$$d(r) \geq \max\{d_{\max}(1-r)^+, \sup_{S\in\{2,3,..,n_{\min}\}} S\ d_{(n_0,n_1^S,\ldots,n_L^S,n_{L+1})}(r)\}. \quad (39)$$

The justification for retaining the term $d_{\max}(1-r)^+$ as part of the maximization is because the original network is fully-connected and by adopting the matching-forward-directed-paths strategy even in the presence of multiple antennas at the source and sink (see Theorem 4.6) we can achieve the same lower bound for half-duplex networks as well.

Example 2: For the case of $(2,4,2)$ network with half-duplex constraint, the proposed protocol achieves the same DMT as the full-duplex case of *Example 1*. However, the FF protocol used naively for a half-duplex system by activating alternative layers during alternate time slots will entail multiplexing-gain loss by a factor of $\frac{1}{2}$.

C. KPP(I) Networks

We consider KPP(I) networks with all nodes, including the source and the sink, having multiple antennas.

1) Full-Duplex KPP(I) Networks: We consider full-duplex KPP(I) networks with multiple-antenna nodes. In the case of single-antenna KPP(I) networks, we activated all backbone paths for equal durations of time in order to obtain a linear DMT in Corollary 2.1. We will use a similar protocol here except that we activate different paths for different durations of time.

Let $\mathbf{H}_{ij}$ be the fading matrix on edge e_{ij}. Let the product fading matrix along backbone path P_i be $\mathbf{G}_i$. Then $\mathbf{G}_i = \prod_{j=1}^{n_i} \mathbf{H}_{ij}$. Let the DMT corresponding to this product matrix $\mathbf{G}_i$ be $d_i(r)$, which can be computed according to formulae given in [9].

Since activating different paths can potentially have different DMTs, it is not optimal in general to use all paths equally. When one is operating at a higher multiplexing gain, one might want to use a path with higher multiplexing gain more frequently in order to get greater average rate. While operating at a low rate, all the paths must be used in order to get maximum diversity. We consider a generic case where path i is activated for a fraction f_i of the duration. [7] These fractions can be chosen depending on r in order to maximize $d(r)$.

By so doing, we will get a parallel channel with repeated coefficients. The DMT of such a channel was evaluated in Lemma 1.5. After making suitable rate adjustments, we obtain a lower bound on the DMT of the protocol as,

$$d(r) \geq \sup_{(f_1,f_2,\cdots,f_K)} \inf_{\left\{\begin{array}{c}(r_1,r_2,\cdots,r_K):\\ \sum_{i=1}^K f_i r_i = r\end{array}\right\}} \sum_{i=1}^K d_i(r_i) \quad (40)$$

2) Half-Duplex KPP(I) Networks: From Section III, we know that under the half-duplex constraint, there exists a protocol activating the K paths equally for KPP(I) networks with $K \geq 3$ causing only causal interference. We can use the same protocol notwithstanding the fact that the relays contain multiple antennas. By doing so, we will get a transfer matrix which will be blt. Also, the diagonal entries of this channel matrix would remain the same as though the relay nodes operate under full-duplex mode. By Theorem 1.3, this gives a lower bound on the DMT, and it is equal to DMT lower bound of the full-duplex network in (40). However the protocol for KPP(I) networks for $K = 3$ activates all paths for equal fractions of time, which is equivalent to setting $f_i = \frac{1}{3}$. Therefore even when there is half-duplex constraint, we can achieve the same DMT given by the (40) with $f_i = \frac{1}{3}$ instead of maximization over all possible f_i.

If we want to achieve different fractions of activation for different parallel paths, then we can follow a different trick for $K \geq 4$. In this case, we can use the $\binom{K}{3}$ 3-parallel path networks, but activate each 3PP network for a different fraction of time, employing the same equi-activation protocol as described in Section III. Hence a path within a 3PP network is activated one-third fraction of the duration for which the 3PP is activated. Thus we use a $3PP$ network as a fundamental unit in the strategy, and for that reason, any fraction of time of activation, f_i for a particular path P_i is limited by $\frac{1}{3}$. In many cases, $f_i > \frac{1}{3}$ may turn out to be infeasible. Moreover, one can show that, for $K \geq 4$, all time fractions $f_i,\ 1 \leq i \leq K$ are feasible as long as $(f_1, f_2, ..., f_K) \in \mathcal{F}$ where

$$\mathcal{F} := \{(f_1, f_2, ..., f_K) : \sum_{i=1}^K f_i = 1, \quad 0 \leq f_i \leq \frac{1}{3}\}.$$

This is shown in Appendix IV.

For $K \geq 4$, this yields a DMT of

$$d(r) \geq \sup_{(f_1,f_2,\cdots,f_K)\in\mathcal{F}} \inf_{\left\{\begin{array}{c}(r_1,r_2,\cdots,r_K):\\ \sum_{i=1}^K f_i r_i = r\end{array}\right\}} \sum_{i=1}^K d_i(r_i) \quad (41)$$

This is the same as the lower bound on the DMT for the full-duplex case, except that we are constrained to have all activation fractions f_i to be lesser than one-third.

VI. Code Design

A. Design of DMT-Achieving Codes

Consider any network and AF protocol described above, and let us say the network is operated for N slots using such a protocol to obtain an induced channel matrix $\mathbf{Y} = \mathbf{HX} + \mathbf{W}$

[7] A similar technique can be used for full-duplex fc layered networks with multiple antennas to improve the achievable DMT.

where X is a $(M \times 1)$ vector and $\mathbf{Y}, \mathbf{W}$ are $P \times 1$ vectors and $\mathbf{H}$ is a $P \times M$ matrix. However, to achieve the DMT of this induced channel one needs to code over both space and time, i.e, transmit a matrix $\mathbf{X}$ drawn from a space-time (ST) code $\mathcal{X}$ as opposed to just sending a vector. In order to obtain an induced channel with $\mathbf{X}$ being a $M \times T$ matrix, we do the following: Instead of transmitting a single symbol, each node transmits a row vector comprising of T symbols during each activation. Then the induced channel matrix takes the form: $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}$, where now, $\mathbf{X}$ is a $M \times T$ matrix, $\mathbf{Y}, \mathbf{W}$ are $M \times T$ matrices with channel matrix H remaining as before. We will regard the product MT as representing the block length of the ST code $\mathcal{X}$ since the transmission of code matrix $\mathbf{X}$ takes place over MT channel uses.

Now from [5], we know that if the code matrix is drawn from an approximately universal code $\mathcal{X}$, then the code $\mathcal{X}$ will achieve the optimal DMT of the channel matrix H irrespective of the statistics of the channel. Explicit minimal delay approximately universal codes for the case when $T = M$ are given in [6], constructed based on appropriate cyclic division algebras (CDA) [11]. These codes can be used here to achieve the optimal DMT of the induced channel matrix.

B. Short DMT-optimal Code Design for Block-Diagonal Channels

In the special case that the channel matrix H is a block-diagonal matrix, we can use MIMO parallel channel codes to construct DMT optimal codes of shorter block length, thereby entailing lesser decoding complexity and delay.[8] In particular if $\mathbf{H}$ is block diagonal with entries $\mathbf{H}_i, i = 1, 2, ..., L$ on the diagonal, with H_i of size $m_i \times p_i$. Since H is of size $P \times M$, we have $M = \sum m_i, P = \sum p_i$. Let $T := \max_i m_i$. Let us construct the code X of size $M \times T$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_L \end{bmatrix}, \tag{42}$$

where $\mathbf{X}_i$ is a $m_i \times m$ matrix, with $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_M$ forming an approximately universal MIMO parallel channel code, then the constructed code is DMT optimal. CDA-based ST codes construction for the rayleigh parallel channel were provided in [22], [23]. This construction was shown to be approximately universal for the class of MIMO parallel channels in [7]. These codes are DMT optimal for every statistical description of the parallel channel and therefore are DMT optimal in this setting as well. Thus the parallel-channel code $\mathcal{X}$ constructed here will have code matrices of size $M \times (\max\{m_i\})$ in place of the earlier size $M \times (\sum m_i)$.

Example 1: (MIMO-NAF) Consider the MIMO NAF protocol for the N relay channel introduced by [8] and explained in *Example-4* of section $II.E$ in [1]. This protocol in essence, is the sequential concatenation of N protocols for a single-relay NAF channel, each involving a different relay. If the source has n_s antennas and destination n_d antennas, the channel matrix will then be block diagonal with each block being of size $2n_d \times 2n_s$, i.e., $m_i := 2n_s, i = 1, 2, ..., N$. Therefore the code design for this block diagonal H will result in a DMT-optimal code of size $2Nn_s \times 2n_s$, which gives essentially the same code as in [8]. In particular for the NAF protocol with a single antenna at all nodes, this yields a code of size $2N \times 2$. Now, even if the paths are activated for unequal fractions of time as suggested in the latter part of *Example-4*, the matrix remains block-diagonal and thus this short-code construction technique can be used here as well.

Example 2: KPP Networks For KPP networks, the matrix between input and the output is a diagonal matrix of size $nK \times nK$ with the diagonal entries comprising of the K product coefficients $\mathbf{g}_i$ repeated periodically. If we consider the $K \times K$ sub-matrix $\mathbf{H}_{\text{sub}}$ of $\mathbf{H}$ comprising of all the K product coefficients the DMT is still the same, i.e., the DMT is equal to $K(1-r)^+$. So we will restrict our attention to this $K \times K$ sub-matrix $\mathbf{H}_{\text{sub}}$. Since this matrix is diagonal, the code construction for block-diagonal channel above gives a parallel channel code of size $K \times 1$ that will be DMT optimal for this channel. For KPP networks with multiple antennas, the same technique will yield a code of size $Kn_s \times n_s$ where n_s is the number of antennas at the source.

C. Short DMT-optimal Code Design for KPP(I) Networks

For KPP(I) networks the matrix is not block-diagonal and therefore the block diagonal code construction can not be used and the longer code construction affords a code length of $M \times M$. Also we need M very large for the initial delay overhead to be minimal. This entails a very large block length, and indeed very high decoding complexity. Now a natural question is whether optimal DMT performance can be achieved with shorter block lengths. We answer this question for KPP(I) networks by constructing DMT optimal codes that have $T = L$ and a block length of L^2, where L is the period of the protocol used. We also provide a DMT optimal decoding strategy that also requires only decoding a $L \times L$ space time code at a time. This is a constant which does not depend on M and therefore, even if we make M large, the delay and decoding complexity are unaffected.

Consider the first L inputs $x_1, x_2, ..., x_L$. If the channel matrix is restricted to these L time slots alone, then channel matrix would be a lower triangular matrix with the L independent coefficients g_i, $i = 1, 2, .., K$ repeated periodically. The DMT of this matrix, after adjusting for rate, is $d_K(r) = K(1-r)^+$. So if we use a $L \times L$ DMT optimal matrix as the input (this can be done by setting $T = L$ and using a $L \times L$ approximately universal CDA based code for the input), we will be able to obtain a DMT of $d_K(r)$ for this subset of the data. This means that the probability of error for this vector comprising of T input symbols will be of exponential order $P_e \doteq \rho^{-d_K(r)}$ if an ML decoder is used to decode the $L \times L$ matrix. Then we cancel this portion of input and then focus on the next L inputs. Again the transfer matrix between input and output will be lower triangular with the same properties, yielding a DMT of $K(1-r)^+$. However, the probability of first block error increases the net probability of decoding error

[8]The same technique can be used when, after permuting the inputs or the outputs, the matrix H is block diagonal.

for the second block by a factor of two. Since this constant factor does not matter in the scale of interest, we conclude that the DMT achieved remains as $K(1-r)^{+}$. This can be repeated for the whole matrix in a successive manner. Thus, the above mentioned successive-interference-cancelation (SIC) based technique yields an optimal DMT while reducing the decoding complexity significantly.

D. Universal Full-Diversity Codes

Consider a input output equation of the form $\mathbf{Y}=\mathbf{H}\mathbf{X}+\mathbf{W}$ where $\mathbf{X},\mathbf{Y},\mathbf{H},\mathbf{W}$ are $M\times M$ matrices.

Usually the code design criterion given for a input matrix to have full diversity for rayleigh fading is that the difference of any two possible input matrices be of full rank. In this section, we show that such a criterion is sufficient to get full diversity under *any* statistical description of the channel matrix. By full diversity here, we mean that the code will attain a diversity equal to $d(0)$ for the corresponding channel.

We quote the following theorem from the theory of approximately universal codes (Theorem 3.1 in [5]):

Theorem 6.1: [5] A sequence of codes of rate $R(\rho):=r\log\rho$ bits/symbol is approximately universal over the MIMO channel if and only if, for every pair of codewords,

$$\lambda_1^2\lambda_2^2\cdots\lambda_{n_{\min}}^2 \geq \frac{1}{2^{R(\rho)+o(\log\rho)}} = \frac{1}{\rho^r\ 2^{o(\log\rho)}}, \tag{43}$$

where $\lambda_1,\ldots,\lambda_{n_{\min}}$ are the smallest $n_{\min}$ singular values of the normalized (by $\frac{1}{\sqrt{\rho}}$) codeword difference matrix. A sequence of codes achieves the DMT of any channel matrix if and only if it is approximately universal.

For the particular case of zero multiplexing gain in Theorem 6.1, we obtain that the criterion reduces to

$$\lambda_1^2\lambda_2^2\cdots\lambda_{n_{\min}}^2 \geq \frac{1}{2^{o(\log\rho)}}. \tag{44}$$

As a result, if for all pairs of codewords, a code satisfies the condition that difference determinant is non-zero, i.e.,

$$\lambda_1^2\lambda_2^2\cdots\lambda_{n_{\min}}^2 \geq L > 0, \tag{45}$$

then that code is approximately universal at rate $r=0$, and thus will achieve, the maximum possible diversity gain $d(0)$ of any given channel matrix.

This criterion is the same as the criterion for full diversity on a Rayleigh channel. This means that all codes with full diversity designed for the rayleigh fading MIMO channel are indeed full diversity for MIMO channels having an arbitrary fading distribution. This is summarized in the following lemma:

Lemma 6.2: A code having non-zero difference determinant achieves full diversity $d(0)$ over any (arbitrary) fading channel.

Therefore we can use a full-diversity code designed for a rayleigh fading MIMO channel to get full-diversity for any KPP or Layered network, when used along with the corresponding protocol for these networks.

Appendix I
Proof of Lemma 3.4

We begin with some definitions.

Definition 13: A *partition* is defined as a set of K nodes obtained by selecting precisely one node from each of the K parallel paths.

We use the term partition here in the sense of a boundary separating one part of the graph from the other, although we do permit edges to cross the partition. The nodes on each backbone path can be assumed to be ordered from left to right. Therefore it is meaningful to speak of nodes that are to the left of the partition and nodes that are to the right of the partition. We use this natural partial order on the set of nodes to define a partial order on partitions.

Definition 14: A partition $\mathcal{P}_1$ is said to be on the left of another partition $\mathcal{P}_2$ if on each parallel path, if the node corresponding to partition $\mathcal{P}_1$ is to the left of, or the same as, the node corresponding to partition $\mathcal{P}_2$.

Definition 15: Given a subset $S\subseteq\{1,2,..,K\}$, a KPP(I) network restricted to S is defined as the subgraph consisting of nodes present in the $|S|$ parallel paths specified by the set S.

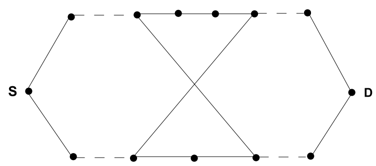
Definition 16: (*k-Switch*) Consider a subset $S\subseteq\{1,2,..,K\}$ of size $|S|=:k>1$, and a KPP(I) network restricted to the subset S. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two partitions, such that $\mathcal{P}_1$ is to the left of $\mathcal{P}_2$. Construct a bipartite graph where the vertices on the left side of the bipartite graph are identified with the left nodes of the partition and those on the right are associated with nodes on the right side of the partition. Thus there are a total of $2\mid S\mid$ nodes in the bipartite graph. We draw an edge in the bipartite graph, between a node on the left and a node on the right if the corresponding nodes are connected by an edge which does not lie on any backbone path. However, if partitions share a node on a particular backbone path, we do not draw an edge between them. If the subgraph comprising the nodes in S on the left and right has a complete matching, then the two partitions are said to form a k-switch corresponding to the set S.

Definition 17: A k-switch corresponding to the set S is said to be *contiguous* if there are no nodes lying in between the two partitions on any of the k backbone paths in the network restricted to the set S. Otherwise, the k-switch is said to be *non-contiguous*.

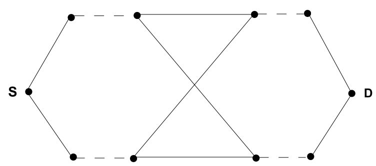
Examples of a 2-switch and a 3-switch in a 3PP network are shown in the figures Fig. 14 and Fig. 15.

Having the necessary definitions in place, we now proceed to specify a protocol for a $3PP$ network in terms of delays introduced at nodes as well as a schedule of edge activations. It is appropriate to recall here that a KPP(I) network can be viewed as a union of a backbone KPP network along with links that connect between various nodes in the network, which we will term as interference links. The particular schedule advocated here is described below:

1) Step 1 - Preprocessing: We begin by deactivating nodes in backbone paths that are encompassed by a shortcut (see Fig.) and redefine the backbone paths to include the shortcut. By deactivating a node, we mean that we will request the node to never transmit. Since there are no transmissions from that node, it can be effectively deleted from the graph. This will lead to a graph in which no backbone path possesses a

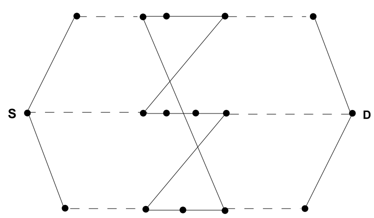


(a) Non-contiguous 2-switch

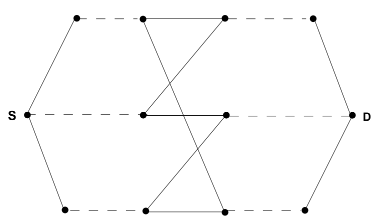


(b) Contiguous 2-switch

Fig. 14. Examples of 2-switches in KPP(I) networks



(a) Non-contiguous 3-switch



(b) Contiguous 3-switch

Fig. 15. Examples of 3-switches in KPP(I) networks

shortcut. This means that all interference links will connect nodes in two distinct back-bone paths. Next, we proceed to remove all the non-contiguous 2-switches, and 3-switches as explained in the lemma below.

Lemma 1.1: Any KPP(I) network can be converted to a second KPP(I) network which does not have any non-contiguous k-switches, for $2 \leq k \leq K$, by deactivating certain nodes from the network and appropriately redefining the backbone paths.

Proof: Let us consider the given KPP(I) network. We employ the following algorithm on the KPP(I) network to perform this conversion:

1) Identify any non-contiguous k-switch in the network for any $2 \leq k \leq K$. If there are no non-contiguous k-switches in the network, terminate the algorithm.
2) Given a non-contiguous k-switch, re-define the backbone path in such a way that the segment of the newly-defined backbone paths lying in between the left and right partition are precisely the edges corresponding to the complete matching. After redrawing the KPP network, it can be seen that there are nodes which are a part of none of the backbone paths. Deactivate these nodes. Since the k-switch was non-contiguous, there exists at least one such node. It can be readily verified that this reconfiguration does not affect the number of backbone paths.
3) Repeat Step 1).

It remains to verify that the algorithm terminates. This is clearly the case, since the algorithm deactivates at least one node during each iteration and this process can not go on indefinitely because there are only a finite number of nodes. (It is for this reason, that this procedure excludes contiguous switches where there will be no node to deactivate).

Note that throughout the iterative process, the number of backbone paths has always remained fixed at K. ■

The above lemma establishes that it is indeed possible to remove all non-contiguous switches for any given KPP(I) network.

2) Step 2 - Layering: After preprocessing the network, we decompose the network into sections which we call layers. As we shall see later, this decomposition of network into layers will be very useful in identifying a DMT optimal schedule for the network.

Definition 18: Two partitions, one on left of the other, are said to create a *layer* if there are no links from any node in the left (right) of the layer to a node inside or to the right (left) of the layer, i.e., no links cross either partition.

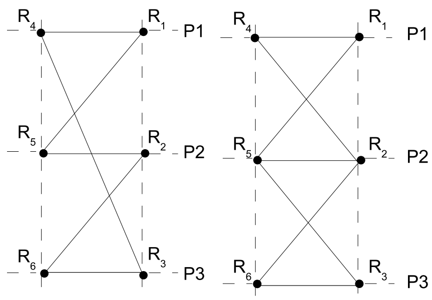
Definition 19: A network is said to be decomposed into a set of layers $L_1, L_2, \ldots, L_N$ if the network can be split into layers L_i such that the right partition of layer L_i is the left partition of layer L_{i+1}.

Now, we attempt to decompose a given KPP(I) network into a set of layers with each layer having certain properties that will be useful for us to obtain an efficient schedule for the network. We note that, with the set of nodes that are connected directly (i.e., by a single edge) to the destination as the left partition, and the destination itself as the right partition, a natural layer is formed. We will refer to this as the sink layer. An analogous definition yields the source layer.

Remark 13: The definition of a layer does not preclude the possibility of having interference links connecting nodes within a partition. We will prove later that such interference links will not present non-causal interference. With this foresight, we will neglect the links present on partitions for now and later return to demonstrate that these do not change the causal nature of the interference.

We next proceed to segment the entire network into a series of layers, each following the next. The layering process is sequential in that we will be ready to identify the second category of layer only after we have identified and segregated layers belonging to the first category.

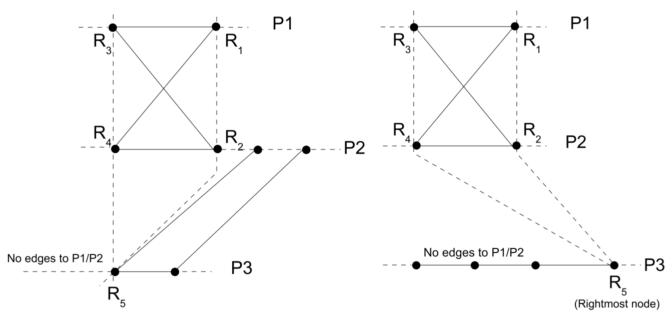
Definition 20: A pair of partitions is said to form a *T-3 layer* (short for layer of Type-3) if the partitions either form a 3-switch which is contiguous or else contains two contiguous 2-switches between different pairs of paths, see Fig. 16 for an example. Apart from layers of this type, we will also regard the source and sink layers as T-3 layers.



(a) Case 1 (b) Case 2

Fig. 16. T-3 layers in KPP(I) network with $K = 3$

It must be noted that it has to be proved that a pair of partitions comprising a contiguous 3-switch or two contiguous 2-switches is indeed a layer. This can be proved by contradiction, by showing that if the pair of partitions is not a layer, then there must exist a non-contiguous 3-switch or a non-contiguous 2-switch in the network, which have been assumed to be removed in the previous preprocessing step.



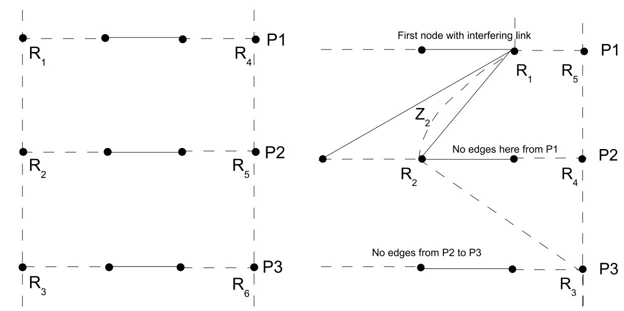
(a) Case 1 (b) Case 2

Fig. 17. Type T-2 layers in a KPP(I) network with $K = 3$

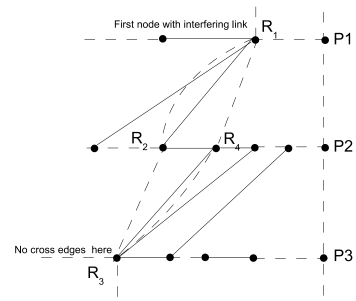
If there exists no T-3 layer in the network, then we proceed to identifying the next type of layer (T-2). Having identified a T-3 layer in the network, we construct a modified graph after removing the internal nodes from the graph. This is for the purpose of identifying the remaining layers in the graph, and does not imply that the internal nodes or edges are deactivated. The obtained network graph will comprise of two disconnected components, which we will call as fragments. We call this process as segregating a layer, which essentially means removing all internal components of a layer. Now we operate in the obtained graph, and try to identify layers in the remaining components of the graph. Once we have identified all the T-3 layers in the graph, we can proceed to decompose the remaining fragments into layers.

Definition 21: Given a fragment of the network, consider a contiguous 2-switch between two paths (say paths P_1 and P_2). Consider the leftmost node in P_3 connected to the right of the 2-switch on paths P_1 or P_2 (if there is no node connected to the right of the 2-switch, consider the rightmost node in P_3 in the fragment). Choose the two nodes in the left partition of the 2-switch along with this node as the left partition and choose the nodes in the right partition of the 2-switch along with the same node in P_3 as the right partition for a layer. These two partitions can be shown to form a layer, using the fact that the fragment does not contain any T-3 layers. We call a layer of this type as a *T-2 layer*, see Fig. 17 for two examples of T-2 layers.

Continuing the sequential layering process, after all T-3 and T-2 layers have been segregated, we are once again left with fragments of the network. Any such fragment does not contain a layer of type T-2 or T-3 and is thus guaranteed not to have any switches. We proceed to decompose these fragments as follows:



(a) Case 1 (b) Case 2



(c) Case 2

Fig. 18. T-1 layers in KPP(I) network with $K = 3$

If there are no interference links in the given fragment, then the whole fragment is treated as a layer. An example of such a layer is given in Fig. 18(a). If there are interference links, we pick a backbone path to which this link is connected and label it as path P_1. Find the rightmost node in P_1, say R_1, that is connected to an interference link. Let the other end of this interference link be connected to a node R_2 in a second backbone path, which we label as path P_2. Conceivably, node R_1 could be connected via other interference links to nodes in path P_2 that lie to the right of node R_2 on path P_2. If this is the case, then we choose the rightmost such node and relabel this node on path P_2 as node R_2. Then

- if there are no interference links connecting to a node on P_2 to the right of R_2, then R_1, R_2 and the rightmost node of the remaining backbone path P_3 in the fragment, form a left partition; this left partition along with the rightmost three nodes of the fragment acting as the right partition, can be verified to be a layer. An example of such a layer is given in Fig. 18(b)
- if there is an interference link connected to a node on path P_2 that is to the right of R_2, this interference link must originate from the remaining backbone path P_3 since the current fragment does not contain any switches; choose

the leftmost node R_3 in P_3 connected to the right of R_2; this node could potentially be connected to multiple nodes in P_2. Let R_4 be the leftmost node in P_2 connected to R_3. Now set R_1, R_2 and R_3 as the three nodes for the left partition and R_1, R_4 and R_3 as the three nodes for the right partition. An example of such a layer is given in Fig. 18(b).

A feature of the layers produced by this last step is that no switches are contained within the layer and we will label such layers as Type-1 layers (T-1). This sequential procedure has resulted in the decomposition of the entire network into layers of type T-3, T-2 or T-1.

Once the network is layered, we assign colors to edges, indicating their activation pattern, as well as delays to nodes.

3) Step 3 - Assigning colors and delays: In this final step, we construct the protocol for the given network by suitably assigning to each edge, a color that represents the activation time slot. We also introduce delays at node in order to make the protocol causal. Note that the given network does not possess any non-contiguous switches and that the network has been layered into N layers, $L_1, L_2, \ldots, L_N$, with each layer being of type T-1, T-2 or T-3.

The conditions corresponding to causal interference (Prop. 4) can be satisfied if at every node where there is an interference link branching out from the backbone path, the shortest delay from the node to the destination through the branch-out link is strictly greater than the delay on the backbone path. It is easily seen that this condition does not depend on the network to the "left" of this node. For this reason, in designing the protocol, we can begin at the rightmost node and make sure that this condition is satisfied for all nodes.

Hence, to construct the protocol, we start from the rightmost layer L_N and proceed toward left layers, assigning colors to edges and delays to nodes in each layer. At the end of this process, we will obtain a protocol for communication, where every node will be given a delay that must be added to every input that it receives and every edge will be given a color (or equivalently, a time slot) to amplify and forward the last symbol that it received. We adopt the convention that a delay of zero at a node corresponds to edges on either side of the node in the backbone path transmitting consecutively. Under this convention, a delay which is not equal to $2 \pmod 3$ can be added to any node without violating the half-duplex constraint at the node.

On each layer L_i, we will make sure that the two criteria below are met. We first provide an outline of the procedure and then explain the details.

- From any node in the layer, we will make sure that the delay to the right partition, along the particular backbone path on which the node is situated, is strictly less than the delay to the right partition on any other path that might lead from the same node (In making this assessment, we will ignore any links connecting nodes within a partition as noted in Note 13). We will ensure this by adding delays to the three nodes in the right partition of the layer, which we will denote by D_i^r, where D_i^r is a three-component vector comprising of the three delays to be added. We refer to this as right-compensation of a layer.
- We will further add delays to the nodes in the left partition D_i^l so that the delays incurred in travelling from left partition to right partition along any of the backbone paths is the same. We refer to this as left-compensation of the layer. It is easy to see that adding delays on the left partition does not change the right compensation of the layer.

For layers of type T-1 it can be easily proved that there exists delays that can lead to right-compensation and left-compensation. Further, if D_i is a right compensation vector for the T-1 layer, there exists another right compensation vector $D_i^{'}$ such that D_i and $D_i^{'}$ differ in only one component and even in that component the difference in the values is equal to one. We will utilize this degree of freedom in choosing the right compensation vector for T-1 layers later.

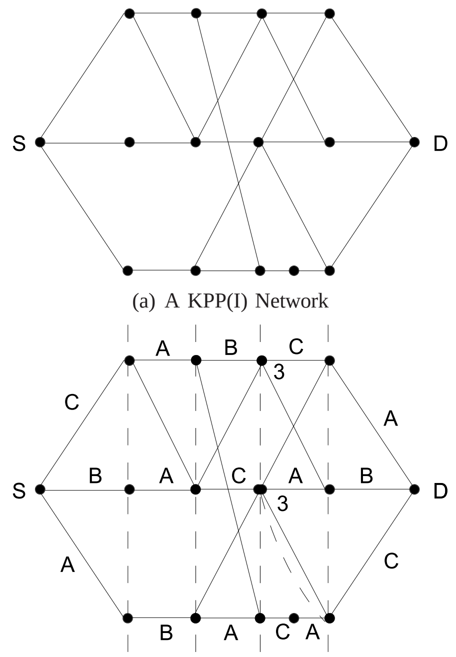
However, as it turns out, just adding delays is not sufficient for layers containing switches, i.e., for layers of type T-3 and T-2. So for these layers, we resort to "neutralizing" all interfering links by carefully designing the protocol. An interference link is considered neutralized under a protocol, if the receiving node of the interference link is scheduled to receive at a different time than the time during which the transmitting node is active. We will explain this process in detail in what follows by detailing how the delays, and colors are assigned within each layer.

The rightmost layer L_N is the destination layer, i.e., it is comprised of three single-edge paths connecting from the backbone paths to the destination. We assign three colors A,B and C to the three paths, so that the transmissions to the destination are all orthogonal. We set D_N^l as the zero vector since no left-compensation is required for this layer. Let $i = N$.

Whenever the layer L_{i+1} is compensated and colors are assigned to the edges inside it, the following information is passed on to the layer L_i that is immediately to its left: its left delays, D_{i+1}^l and the colors on the edges immediately to the right of layer L_i, which we call as right colors and which can be regarded as a vector comprising of the three colors on the three respective, ordered, backbone paths.

If layer L_i is a T-1 layer, the delays D_i^r for right-compensation are computed. The delays $D_i^r + D_{i+1}^l$ are added to the right partition. However, if this gives delays that violate half-duplex constraints (i.e., one of the delays turns out to be equal to 2 modulo 3), then the degree of freedom in the $T1$ layer is utilized to add more delay to a path so that the delays do not violate the half-duplex constraint in any of the paths. Inside the layer, the edges are colored consecutively so that there is no delay at any node inside the layer. The left-compensation delays D_i^l are computed so that the delays on all the backbone paths become equal. We make an additional modification to this layer in the special case that the layer L_{i-1} to the left is a T-3 layer and all the right colors output to layer L_{i-1} are the same. In this case, the degree of freedom in choosing the right compensation delays for the layer L_i is utilized so that the right colors of layer L_{i-1} are not the same. The reason for this modification will become clear later when we consider T-3 layers.

If layer L_i is a T-2 layer, we first color the two edges of



(a) A KPP(I) Network

(b) The network after layering, coloring and delaying

Fig. 19. Adding colors and delay in KPP(I) network

the contiguous 2-switch on the back-bone path with distinct colors. It can be verified easily that this can always be done without violating the half-duplex constraint irrespective of the right colors given to layer L_i. This coloring neutralizes the interference links in the layer and therefore, there is no need for right compensation of the layer, i.e., we set $D_i^r = 0$. However delays still need to be added to the right partition of the layer since this is also the left partition of layer L_{i+1}. Delays are added to the right partition such that the delay on each node on the right partition is greater than or equal to the delays D_{i+1}^l. We compute the left compensation vector D_i^l to ensure that all the backbone paths have equal delays.

If layer L_i is a T-3 layer, all the interference links can be effectively neutralized if the three edges on the backbone paths can be assigned distinct colors. It can be easily shown that this can be done without violating the half-duplex constraint as long as the right colors for this layer are not all the same. As an example, if the right colors are (A, B, B), then we can choose the colors in layer L_i as (B, A, C). However if the right colors are (A, A, A), we cannot get distinct colors on the layer L_i without violating the half-duplex constraint, since no edge in layer L_i is allowed to have color A. Therefore, we must ensure that the right colors are all not the same. The right colors will not all be the same if the layer L_{i+1} is a T-3 or a T-2. If the right layer is a T-1 layer, then it has already been modified in such a way that the right colors of L_i are not all same. Thus all the interference links in the T-3 layer can be neutralized by choosing distinct colors on the three paths. Thus, just as in the T-2 case, there is no need for right compensation of the layer, i.e., we set $D_i^r = 0$. Delays are added to the right partition such that the delay on each node on the right partition is greater than or equal to the delays D_{i+1}^l. We compute the left compensation vector D_i^l to ensure all the backbone paths have equal delays.

Now the delay from any node in layer L_i to the right partition of the layer is strictly greater than that along the corresponding backbone path to the right partition. Once the right partition has been reached, both paths incur equal delay (since this would have been ensured in layer L_{i+1} itself). In this way, we have satisfied the conditions of Prop. 4, hence the interference is causal under the designed protocol. An example of a KPP(I) network is given in Fig. 19(a) and after layering, coloring and delaying the network is depicted in Fig. 19(b). In this figure, the numbers denote the amount of delay to be added at that particular node. For example, the delay of 3 in Fig. 19(b), indicates that the node waits for 3 time slots (which amounts to one protocol cycle).

However, as noted in Remark 13, we have not accounted for the interfering links between nodes within a partition while constructing the protocol. In the proposition given below, we justify that interference links connecting nodes inside a partition do not alter the causal nature of the protocol.

Proposition 5: Assume that a protocol is designed for a KPP(I) network by the procedure detailed above. Then the edges that are present within any partition of the network do not alter the causal nature of the protocol.

Proof: By hypothesis, the protocol is constructed for a KPP(I) network by the procedure detailed above. Hence, while checking for the condition that a given interference link causes purely causal interference, links inside a partition are discounted. We need to show that notwithstanding this fact, the links inside a partition do not result in any non-causal interference. Equivalently, it suffices to show that for any link inside a left partition, the delay along the backbone path leading from the tail of the interference link to the right partition is lesser than that along any other path to the right partition from the same node.

Let the network be decomposed into N layers, $L_1, L_2, \ldots, L_N$. Without loss of generality, consider an interference link in the left partition $\mathcal{P}_i$ of an arbitrary layer $L_j,\ 1 \leq j \leq N$ connecting nodes R_1 and R_2 on the first and second backbone paths, say P_1 and P_2. Then focussing on the path from R_1, we will prove that the link under consideration does not create any non-causal interference. By symmetry the same will hold even if we choose R_2. Assume that delay encountered by a symbol forwarded via the backbone paths P_1 and P_2 within the layer L_j be d_1 and d_2 respectively. As part of left compensation, let $d_1^{'}$ and $d_2^{'}$ be added to nodes R_1 and R_2 respectively.

Any symbol which is received at R_1 is delayed for $d_1^{'}$, and then forwarded. Then the symbol simultaneously starts following both the backbone path P_1 as well as the path via the interference link. Therefore the delay $d_1^{'}$ does not come into picture for this relative delay comparison and we have that the delay through the shortcut path that includes the interference link is equal to $1 + d_2^{'} + d_2$ and the delay through the back-bone path from R_1 is d_1. Therefore in order to show that the delay on the back-bone path is strictly lesser than the delay on the short-cut path, we need to show that $d_2^{'} + d_2 \geq d_1$. This is indeed ensured by the algorithm designed above.

Thus the interference remains causal considering the path beginning from the relay R_1. Thus the protocol remains causal even in the presence of interference link within a partition.

■

This completes the proof of the lemma.

Appendix II
Proof of Lemma 4.5

Proof: Let us assume without loss of generality that $N_1 \geq N_2 \geq ...N_{L+1}$. We also have,

$$\mathbf{H} = \begin{bmatrix} \mathbf{g}_1 & & \\ & \ddots & \\ & & \mathbf{g}_N \end{bmatrix}. \tag{46}$$

Consider a variable transformation where $\boldsymbol{\alpha}_j^{(k)}$ is defined such that $\rho^{-\boldsymbol{\alpha}_j^{(k)}} = |\mathbf{h}_j^{(k)}|^2$. Now the DMT $d(r)$ of the parallel channel is characterized as,

$$\begin{aligned}
\rho^{-d(r)} &\doteq \Pr\{\log\det(I+\rho\mathbf{H}\mathbf{H}^\dagger) \leq r\log\rho\} \\
&= \Pr\{\det(I+\rho\mathbf{H}\mathbf{H}^\dagger) \leq \rho^r\} \\
&= \Pr\{\prod_{i=1}^{N}(1+\rho|\mathbf{g}_i|^2) \leq \rho^r\} \\
&= \Pr\{\prod_{i=1}^{N}(1+\rho\prod_{k=1}^{L+1}|\mathbf{h}_{e(i,k)}^{(k)}|^2) \leq \rho^r\}
\end{aligned}$$

$$\begin{aligned}
&= \Pr\{\prod_{i=1}^{N}(1+\rho\prod_{k=1}^{L+1}\rho^{-\boldsymbol{\alpha}_{e(i,k)}^{(k)}}) \leq \rho^r\} \\
&= \Pr\{\prod_{i=1}^{N}(1+\rho^{1-\sum_{k=1}^{L+1}\boldsymbol{\alpha}_{e(i,k)}^{(k)}}) \leq \rho^r\}
\end{aligned}$$

$$\begin{aligned}
&\doteq \Pr\{\prod_{i=1}^{N}\rho^{(1-\sum_{k=1}^{L+1}\boldsymbol{\alpha}_{e(i,k)}^{(k)})^+} \leq \rho^r\} \\
&= \Pr\{\sum_{i=1}^{N}(1-\sum_{k=1}^{L+1}\boldsymbol{\alpha}_{e(i,k)}^{(k)})^+ \leq r\} \qquad (47) \\
&\leq \Pr\{\sum_{i=1}^{N}(1-\sum_{k=1}^{L+1}\boldsymbol{\alpha}_{e(i,k)}^{(k)}) \leq r\} \qquad (48) \\
&= \Pr\{N-\sum_{k=1}^{L+1}N_k\sum_{j=1}^{M_k}\boldsymbol{\alpha}_j^{(k)} \leq r\}.
\end{aligned}$$

The last equality follows since each $|\mathbf{h}_{ij}|^2$ appear in N_i of the terms in $\mathcal{H}$ irrespective of j and so do the corresponding $\boldsymbol{\alpha}_{ij}$. Let $d_1(r)$ be defined as the SNR exponent of the RHS in the last equation above, i.e.,

$$\Pr\{N-\sum_{k=1}^{L+1}N_k\sum_{j=1}^{M_k}\boldsymbol{\alpha}_{kj} \leq r\} \doteq \rho^{-d_1(r)} \tag{49}$$

Let the set S and T be defined as

$$\begin{aligned}
S &:= \{(\alpha_j^k) : \sum_{i=1}^{N}(1-\sum_{k=1}^{L+1}\alpha_{e(i,k)}^{(k)})^+ \leq r\}, \qquad (50) \\
T &:= \{(\alpha_j^k) : \sum_{i=1}^{N}(1-\sum_{k=1}^{L+1}\alpha_{e(i,k)}^{(k)}) \leq r\}. \qquad (51)
\end{aligned}$$

Now,

$$d(r) = \inf_{(\alpha_j^k)\in S}\sum_{k,j}\alpha_j^k, \tag{52}$$

and

$$d_1(r) = \inf_{(\alpha_j^k)\in T}\sum_{k,j}\alpha_j^k. \tag{53}$$

Since $S \subseteq T$, we have that $d(r) \geq d_1(r)$.

Remark 14: While the minimizing solution (α_j^k) for (53) is usually a member of the set T, if the solution happens to be an element of the set S also, then it follows that $d(r) = d_1(r)$. As will be seen later, this is the case here. We proceed to compute $d_1(r)$ and the optimizing (α_j^k) for (53).

Now,

$$d_1(r) = \inf_{\left\{\begin{array}{c} N-\sum_{k=1}^{L+1}N_k\sum_{j=1}^{M_k}\alpha_j^k \leq r \ , \\ \alpha_j^k \geq 0 \end{array}\right\}} \sum_{k=1}^{L+1}\sum_{j=1}^{M_k}\alpha_j^k. \tag{54}$$

Define $\alpha^{(k)} := \sum_{j=1}^{M_k}\alpha_j^k$ to obtain,

$$d_1(r) = \inf_{\{N-\sum_{k=1}^{L+1}N_k\alpha^{(k)}\leq r \ , \ \alpha^{(k)}\geq 0\}}\sum_{k=1}^{L+1}\alpha^{(k)}. \tag{55}$$

The infimum is attained in the above minimization by $\alpha^{(1)} = \frac{N-r}{N_1}$, $\alpha^{(i)} = 0, \forall i = 2, ..., N$ and the value of the infimum is $\frac{N-r}{N_1}$. Thus $d(r) \geq d_1(r) = \frac{N-r}{N_1}$.

Now we will verify that this lower bound is in fact equal to the DMT of the channel. An optimizing assignment of α_j^k for obtaining $d_1(r)$ is given by, $(\alpha_j^1)^* = \frac{N-r}{N_1M_1} = \frac{N-r}{N}, j = 1, 2, .., M_1$. Clearly $(\alpha_j^1)^* \in T$. It can be easily checked that $(\alpha_j^1)^* \in S$ also. This implies, by Remark 14, that this is the optimizing α_j^k for (52) too. Therefore, we have

$$d(r) = d_1(r) = \frac{N-r}{N_1} \tag{56}$$

■

Appendix III
Achievability of Outage Exponent

Consider a compound channel, where a channel, s is chosen from a set of possible channels $\mathcal{S}$ and the channel once chosen, remains fixed. Then, the compound channel coding theorem tells us that,, for a fixed $p_X(x)$, any rate

$$R < \inf_{s\in(S)} I(\mathbf{X};\mathbf{Y}|\mathbf{S}=s) \tag{57}$$

is achievable on the compound channel, i.e, irrespective of the particular channel s chosen, the probability of error at the

receiver can be reduced to zero.

We will now consider a specific channel $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}$, where the channel $\mathbf{H}$ can be any matrix in $\mathbb{C}^{m \times n}$. However, if we consider the set of all possible $\mathbf{H}$ that are not in outage as a set $\mathcal{H}$, we can apply a compound channel theorem to this set and whenever a matrix is chosen from this set, we can drive the probability of error to zero.

Consider the set of all channels not in outage, $\mathcal{H}$ as the possible channels in the compound channel. Then $\mathcal{H}$ is defined as

$$\mathcal{H} = \{\mathbf{H} : \log\det(I + \mathbf{H}\Sigma_x\mathbf{H}^\dagger > r\log(\rho)\}. \tag{58}$$

Now, any rate lesser than R is achievable on the compound channel by using an optimal compound channel code. This means that there exists a code for this compound channel, whose probability of error is less than ϵ for any given $\epsilon > 0$.

In this setting, let the optimal covariance matrix for minimizing outage probability for a given rate be p_X^*, i.e., $p_X^* = \mathbb{CN}(0, \Sigma_x)$, where Σ_x be optimizing covariance matrix for the following optimization:

$$P_{\text{out}}(R) = \inf_{\Sigma_x \geq 0,\ \text{Tr}(\Sigma_x) \leq n\rho} \Pr(\log\det(I + \mathbf{H}\Sigma_x\mathbf{H}^\dagger \leq nR).$$

If we use this optimal compound channel code on the slow fading channel, we know that the probability of error goes to ϵ whenever, the channel realization $H \in \mathcal{H}$.

The probability of error of this code when used on the slow fading channel is given by

$$\begin{aligned} P_e &= P_{\text{out}}P_{\text{e/out}} + P_{\text{out}^c}P_{e/out^c} && (59)\\ &\leq P_{\text{out}} + P_{e/out^c} && (60)\\ &\leq P_{\text{out}} + \epsilon && (61)\\ &\dot{\leq} P_{\text{out}} && (62)\end{aligned}$$

where P_{out} is the probability of the channel being in outage and P_{out^c} is the probability of the channel not being in outage. $P_{\text{e/out}}$ is the probability of error of the code given the channel is in outage and P_{e/out^c} is the probability of error of the code given the channel is not in outage. Thus the outage probability is achievable.

Appendix IV
Different Fractions of Activation in Multi-antenna KPP(I) Networks

In this appendix, we study the achievable region when different paths are to be activated for different fractions of time. For choosing 3 parallel paths from the KPP network, the total number of possibilities is $\binom{M:=K}{3}$, let us number these possibilities as $1, 2, ..., M$. If we use the 3 paths specified by the combination for a fraction λ_i fraction of time. Let us construct a matrix of size $K \times M$ with each column being composed of distinct vectors of weight 3. Now $\frac{1}{3}A\lambda$ yields a vector $f = [f_1, f_2, ..., f_K]^t$ of size K that gives us the fraction of duration f_i for which the parallel path P_i is activated. Now we have to identify which are the possible fractions f that can be obtained by choosing various combinations of λ. The lemma below proves that all valid activation fractions f such that each component is lesser than $\frac{1}{3}$ can be obtained using this scheme.

Lemma 4.1: Let $K \in \mathbb{Z}$, $K \geq 4$, and let $M = \binom{K}{3}$. Construct a matrix $A \in \{0,1\}^{K \times M}$ constituting of distinct columns, each being a vector of weight 3. Then the equation $\frac{1}{3}A\ [y_1\ \ y_2 \ldots\ \ y_M]^t = [f_1\ \ f_2 \ldots f_K]^t$ has a solution $\lambda = [\lambda_1\ \lambda_2 \ldots \lambda_M]$ satisfying,

1) $0 \leq \lambda_i \leq 1,\ \forall\ 1 \leq i \leq M$ and
2) $\lambda_1 + \lambda_2 + \ldots + \lambda_M = 1$,

for every $[f_1\ \ f_2 \ldots f_K]$ in the region

$$\mathcal{F} = \{[f_1\ f_2\ \ldots\ f_K]^t : \sum_{i=1}^{K} f_i = 1,\ \ 0 \leq f_i \leq \frac{1}{3}\}.$$

Proof: The region $\mathcal{F}$ is convex, and thus every point in $\mathcal{F}$ can be expressed as a linear combination of its extreme points, where the coefficients appearing in the linear combination lie between 0 and 1, and add up to 1. Hence, it is sufficient to prove that the extreme points of $\mathcal{F}$ are $\frac{1}{3}$ times columns of A. The claim is that this is precisely the case, i.e., extreme points of $\mathcal{F}$ are vectors containing $\frac{1}{3}$ as entries in 3 positions, and zero elsewhere.

Suppose it is not the case. Then there exists an extreme point $x = [x_1\ x_2\ \ldots\ x_K]$, such that at least one entry, say x_1 without loss of generality, is less than $\frac{1}{3}$. But, due to constraints of the region, this forces one more entry, say x_2 without loss of generality, to be greater than $\frac{1}{3}$. Then clearly, $\exists\ \delta > 0$ such that

$$\begin{aligned} x_1 + \delta &\leq \frac{1}{3}\\ x_2 - \delta &\geq 0\\ x_1 - \delta &\geq 0\\ x_2 + \delta &\leq \frac{1}{3},\end{aligned}$$

so that $x' = [x_1 + \delta\ x_2 - \delta\ \ldots\ x_K]$ and $x'' = [x_1 - \delta\ x_2 + \delta\ \ldots\ x_K]$ belong to the region $\mathcal{F}$. Now,

$$x = \frac{1}{2}x' + \frac{1}{2}x'',$$

which contradicts our hypothesis that x is an extreme point. This completes the proof. ■

Acknowledgment

Thanks are due to K. Vinodh and M. Anand for useful discussions.

References

[1] K. Sreeram, S. Birenjith, and P. V. Kumar, "DMT of multi-hop cooperative networks - Part I: Basic results," submitted to *IEEE Trans. Inform. Theory*.

[2] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity–multiplexing tradeoff in half–duplex cooperative channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[3] M. Yuksel and E. Erkip, "Multiple–antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective", *IEEE Trans. Inform. Theory* , vol 53, no.10, pp. 3371-3393, Oct. 2007.

[4] P. Elia, K. Vinodh, M. Anand, and P. V. Kumar, "D-MG tradeoff and optimal codes for a class of AF and DF cooperative communication protocols," *IEEE Trans. Inform. Theory,* accepted for publication pending revision. Available Online: http://arxiv.org/abs/cs/0611156 , Nov. 2006.

[5] S. Tavildar and P. Viswanath, "Approximately universal codes over slow-fading channels" *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3233–3258, July 2006.

[6] P. Elia, K. Raj Kumar, S. A. Pawar, P. V. Kumar, and H-F. Lu, "Explicit, minimum-delay space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3869–3884, Sept. 2006.

[7] P. Elia and P. V. Kumar, "Approximately–universal space–time codes for the parallel, multi–block and cooperative dynamic–decode–and–forward channels," *Available Online: http://arxiv.org/abs/0706.3502* , June 2007.

[8] S. Yang and J.-C. Belfiore, "Optimal space–time codes for the MIMO amplify-and-forward cooperative channel," *IEEE Trans. Inform. Theory*, vol. 53, Issue 2, pp 647-663, Feb. 2007.

[9] S. Yang and J.-C. Belfiore, "Diversity of MIMO multihop relay channels," submitted to *IEEE Trans. on Inform. Theory*, Available Online: http://arxiv.org/abs/0708.0386, Aug. 2007.

[10] S. Yang and J.-C. Belfiore, "Towards the optimal amplify–and–forward cooperative diversity scheme," *IEEE Trans. Inform. Theory*, vol. 53, Issue 9, pp 3114-3126, Sept. 2007.

[11] B. A. Sethuraman, B. Sundar Rajan, and V. Shashidhar, "Full–diversity, high–rate, space–time block codes from division algebras," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.

[12] K. Sreeram, S. Birenjith, and P. V. Kumar, "Multi–hop cooperative wireless networks: Diversity multiplexing tradeoff and optimal code design," *Proceedings of Information Theory and Applications Workshop, UCSD*, Feb. 2008.

[13] K. Sreeram, S. Birenjith, K. Vinod, M. Anand, and P. V. Kumar "On the throughput, DMT and optimal code construction of the K-parallel-path cooperative wireless fading network," *Proc. 10th Int. Symp. Wireless Personal Multimedia Commun.*, Dec. 2007.

[14] K. Sreeram, S. Birenjith, and P. V. Kumar, "DMT of multi–hop cooperative networks–Part I: K–parallel–path–networks" *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, July 6-11, 2008.

[15] K. Sreeram, S. Birenjith, and P. V. Kumar, "DMT of multi–hop cooperative networks–Part II: Layered and multi–antenna networks," *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, July 6-11, 2008.

[16] K. Sreeram, S. Birenjith, and P. V. Kumar, "Multi–hop cooperative wireless networks: Diversity multiplexing tradeoff and optimal code design," *Available Online* : http://arxiv.org/pdf/0802.1888 , Feb. 2008.

[17] K. Sreeram, S. Birenjith, and P. V. Kumar, "On the throughput, DMT and optimal code construction of the K–parallel–path cooperative wireless fading network," *USC CSI Technical Report*, CSI-2007-06-07, June 2007.

[18] S. Borade, L. Zheng, and R. Gallager, "Amplify and forward in wireless relay networks: Rate, diversity and network size," *IEEE Trans. Inform. Theory*, vol 53, no.10, pp 3302-3318, Oct. 2007.

[19] S. O. Gharan, A. Bayesteh, and A. K. Khandani, "On the diversity-multiplexing tradeoff in multiple-relay network," submitted to *IEEE Trans. Inform. Theory*, April 2008.

[20] A. S. Avestimehr, S. N. Diggavi, and D. Tse, "Wireless network information flow," *Proc. Forty-fifth Allerton Conf. Commun. Contr. Comput.*, Illinois, Sep 2007.

[21] R. Vaze and R. W. Heath Jr., "Maximizing reliability in multi-hop wireless networks with cascaded space-time codes ," *Proc. Inform. Theory and Appln. Workshop, UCSD*, Feb. 2008.

[22] Hsiao-feng Lu, "Explicit construction of multi-block space-time codes that acheive the diversity-multiplexing gain tradeoff," *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, USA , 2006 .

[23] S. Yang, J.-C. Belfiore and G. Rekaya, "Perfect space-time block codes for parallel MIMO channels," *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, USA , 2006.

[24] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless mesh networks with orthogonal channels," *IEEE/ACM Trans. Networking*, Vol 13, No.4, pp 868-880, Aug. 2005.

[25] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd Edition, John Wiley and Sons, New York, 2006.

[26] A. Ribeiro, X. Cai, and G. B. Giannakis, "Symbol error probabilities for general cooperative links," *IEEE Trans. Wireless Comm.* , Vol. 4, No. 3, pp 1264-1273, May 2005.